A TRIDENT SCHOLAR PROJECT REPORT

NO. 458

Protein Engineering: Development of a Metal Ion-Dependent Switch

by

Midshipman 1/C Emily S. Kilen, USN

UNITED STATES NAVAL ACADEMY
ANNAPOLIS, MARYLAND

USNA-1531-2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
05-22-17		

4. TITLE AND SUBTITLE
Protein Engineering: Development of a Metal Ion-Dependent Switch

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kilen, Emily S.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Naval Academy
Annapolis, MD 21402

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Trident Scholar Report no. 458 (2017)

12. DISTRIBUTION / AVAILABILITY STATEMENT

This document has been approved for public release; its distribution is UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Proteins are biopolymers that perform a myriad of functions in living cells. These functions are determined by each protein's three-dimensional structure. Current knowledge of protein folding principles provides a partial understanding of the thermodynamic factors that drive protein structure, folding, and stability, sufficient to allow proteins to be treated as templates for design and engineering. This project used protein engineering to explore protein structure and folding mediated by interactions with metal ions. As a proof of principle, experiments were undertaken that aimed to re-engineer staphylococcal nuclease to contain a metal ion-dependent switch that exhibits a loss of structure in the absence of a specific metal ion but recovers its native fold in the presence of that ion. Spectroscopic methods were used to monitor structural changes between the metal-free and the metal-bound protein. Changes in the protein's amino acid sequence were introduced systematically to create nickel (II) binding sites based on a naturally occurring high-affinity nickel site. The site was comprised of four amino acid side chains that coordinated the metal ion. Several iterations of candidate proteins, each containing a putative nickel binding site comprised of 2-4 residues, and of reference proteins lacking this site were designed, produced, and characterized spectroscopically. From each of the protein iterations, information was obtained about the refolding process, including the effects of steric constraints, protein oligomerization, and protein thermodynamics. Proteins with fewer substitutions were more effective at maintaining their structure due to a reduced thermodynamic penalty.

15. SUBJECT TERMS
Metalloprotein, protein engineering, staphylococcal nuclease, circular dichroism (CD) spectroscopy

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		63	**19b. TELEPHONE NUMBER** *(include area code)*

U.S.N.A. --- Trident Scholar project report; no. 458 (2017)

PROTEIN ENGINEERING: DEVELOPMENT OF A METAL ION-DEPENDENT SWITCH

by

Midshipman 1/C Emily S. Kilen
United States Naval Academy
Annapolis, Maryland

_____
(signature)

Certification of Advisers Approval

Professor Jamie L. Schlessman
Chemistry Department

_____
(signature)

_____
(date)

Professor Carl E. Mungan
Physics Department

_____
(signature)

_____
(date)

Acceptance for the Trident Scholar Committee

Professor Maria J. Schroeder
Associate Director of Midshipman Research

_____
(signature)

_____
(date)

USNA-1531-2

Abstract

Proteins are biopolymers that perform a myriad of functions in living cells. These functions are determined by each protein's three-dimensional structure. Current knowledge of protein folding principles provides a partial understanding of the thermodynamic factors that drive protein structure, folding, and stability, sufficient to allow proteins to be treated as templates for design and engineering. This project used protein engineering to explore protein structure and folding mediated by interactions with metal ions. As a proof of principle, experiments were undertaken that aimed to re-engineer staphylococcal nuclease to contain a metal ion-dependent switch that exhibits a loss of structure in the absence of a specific metal ion but recovers its native fold in the presence of that ion. Spectroscopic methods were used to monitor structural changes between the metal-free and the metal-bound protein. Changes in the protein's amino acid sequence were introduced systematically to create nickel (II) binding sites based on a naturally occurring high-affinity nickel site. The site was comprised of four amino acid side chains that coordinated the metal ion. Several iterations of candidate proteins, each containing a putative nickel binding site comprised of 2-4 residues, and of reference proteins lacking this site were designed, produced, and characterized spectroscopically. From each of the protein iterations, information was obtained about the refolding process, including the effects of steric constraints, protein oligomerization, and protein thermodynamics. Proteins with fewer substitutions were more effective at maintaining their structure due to a reduced thermodynamic penalty.

Keywords: Metalloprotein, protein engineering, staphylococcal nuclease, circular dichroism (CD) spectroscopy

Acknowledgements

This work was supported at USNA by the Office of Naval Research and by the Defense Threat Reduction Agency (DTRA) Service Academy Initiative and at JHU by the National Institutes of Health (GM-065197 to Dr. Bertrand García-Moreno). Initial studies in this project were performed collaboratively with ENS Nic Butler (USNA, '16). The author gratefully acknowledges assistance from Prof. Virginia Smith, Assoc. Prof. Danny Morse, LT Erick Roman Perez and Mr. Matt Schroeder. Crystal structures for SNase Δ+PHS V23H (Protein Data Bank (PDB) ID 4ZUI), Δ+PHS L25H (PDB ID 5C4Z), Δ+PHS L36H (PDB ID 4LAA), and Δ+PHS V66H (PDB ID 5C3W) used in model design were determined in the García-Moreno lab by Dr. Jaime Sorenson, Erika Wheeler, Dr. Aaron Robinson, and Dr. Jamie Schlessman. Preliminary molecular biology techniques were performed at JHU by Dr. Aaron Robinson, Meredith Peck, and Dr. Jamie Schlessman. Dr. Aaron Robinson collected preliminary spectroscopic data at JHU.

Table of Contents

Background

Proteins are biopolymers, macromolecules composed of repeating amino acid subunits, which are produced in living organisms. Proteins adopt a single native structure, which is dependent on environmental factors, based on their amino acid sequence. Amino acids consist of a backbone containing of carboxyl and amine functional groups surrounding a central alpha carbon, to which is attached one of twenty side chains (figure 1). Successive amino acids are linked covalently through amide bonds to form the linear biopolymer. Amino acid side chain identities can be classified into several categories: (1) ionizable groups that are positively charged under conditions of low pH (acidic); (2) ionizable groups that are negatively charged under conditions of high pH (basic); (3) non-ionizable polar groups; (4) nonpolar aliphatic groups; and (5) nonpolar aromatic groups [1]. Spontaneous protein folding in polar solvent, such as water, occurs due to the hydrophobic effect, in which amino acid residues interact with the surrounding solution and each other to yield the most thermodynamically stable structure. The protein folds spontaneously so that the nonpolar side chains are buried in the protein interior where they do not interact directly with the solution, thereby minimizing the number of ordered solvent molecules around the protein. In this arrangement, amino acid side chains with polar and ionizable groups tend to face outward in aqueous solution, which promotes stabilizing interactions between the protein and water [1].

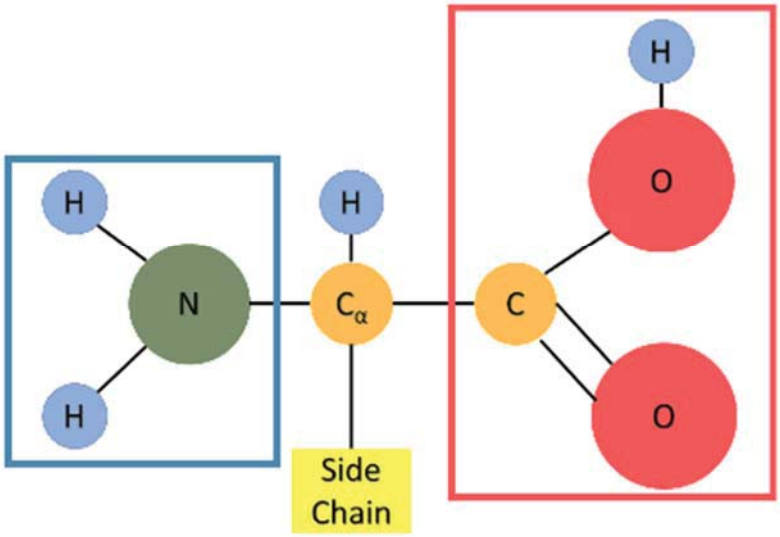

Figure 1: Schematic of amino acid structure depicting the amine functional group, NH_2, (blue box) and the carboxylic acid functional group, COOH (red box), and the variable side chain (yellow) bound to the central alpha carbon, C_α.

Protein structure is described in a hierarchy of four levels (figure 2). Primary structure refers to the protein's amino acid sequence and is maintained by covalent peptide bonds between adjacent amino acid residues. Secondary structure describes local order maintained by main-chain hydrogen bonds. Based on the sequence of amino acids, local folding into one of several types of secondary structural elements may occur: (1) an alpha helix consisting of adjacent residues joined by main-chain hydrogen bonds through a carbonyl oxygen of residue *i* and an amide nitrogen of residue *i*+4, resulting in a spiral appearance; and (2) a beta sheet containing several regions of residues that pack together in parallel or antiparallel arrangements stabilized by main-chain hydrogen bonds. Regions of the protein that do not conform to formal secondary structural elements are referred to as random coil. Secondary structural elements spontaneously pack into a three-dimensional topology that minimizes the energy for the overall protein [1]. The tertiary structure has a single native conformation under physiological conditions and determines the function of the protein. Hydrogen bonding through main-chain and side-chain interactions, electrostatic interactions (ionic and ion-dipole interactions), Van der Waals (induced electrical) interactions, and nonpolar forces all influence tertiary protein structure. Quaternary structure is the assembly of protein subunits to attain a physiologically relevant complex, such as the association of two subunits to form a dimer.

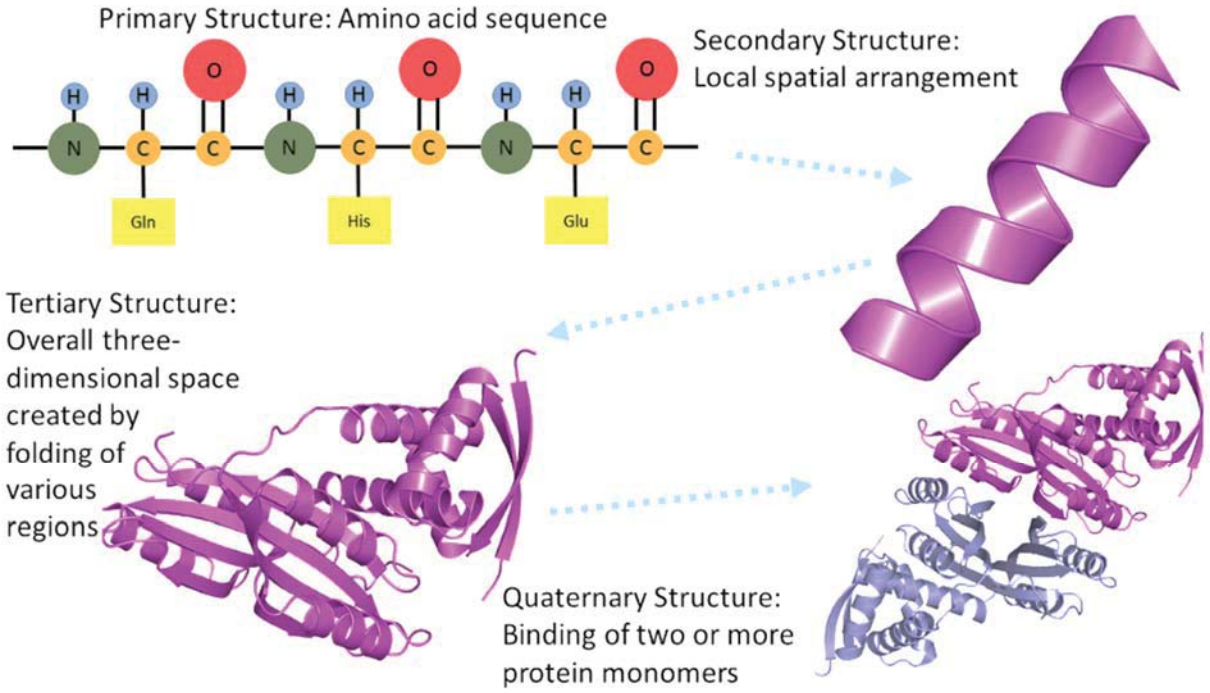

Figure 2: Protein structural hierarchy. Covalently linked amino acids comprise the primary structure, such as the tripeptide sequence shown. Gln, His and Glu refer to amino acids glutamine, histidine, and glutamic acid, an amino acid sequence found in an alpha helix in NikR. Local arrangements of neighboring amino acids form secondary structural elements, such as the alpha helix shown. Tertiary structure is the overall structure of one protein subunit. Assembly of multiple protein subunits comprises quaternary structure, depicted by the subunits of NikR in purple and blue. These figures were generated in Pymol [10]. Because proteins contain too many atoms to display clearly, ribbon diagrams are used to depict the general shape of the protein. These ribbon diagrams are based on experimentally-determined atomic coordinates (2BJ8).

Environmental factors, however, can affect the structure of proteins by influencing their folding, stability, and activity. Solvent polarity, pH, temperature, and chemical denaturants are all capable of inducing changes to protein structure. Another factor is the presence of metal ions, which are required by many proteins. Most naturally-occurring metalloproteins are specific to a particular metal ion, such as ions of magnesium, iron or nickel, and involve direct interactions with specific amino acids in the protein [2]. Secondary and tertiary structural features can change as a protein binds to a metal ion, with potential impact on the protein's ability to function. In addition to their essential roles as catalysts, transport vehicles, and regulatory molecules, metalloproteins are of interest as subjects for protein engineering, in which the protein's amino acid sequence is modified to produce proteins with modified structures and functions [3].

Applications for metalloproteins include their use as dynamic switches capable of conformational changes upon binding with metal ions. These molecules have applications in medicine and nanotechnology [2–3]. Metalloproteins are also of interest because heavy metal ions can be toxic to organisms, as evidenced by the evolution of complex protein machinery in bacteria to detect, sequester, and remove such ions. A potential application of successful engineering of a metal ion-dependent switch protein is as a biosensor to detect specific metal ions. Successful metal ion-dependent switch proteins could provide a basis for the development of other biosensors to detect environmental toxins and chemical threats. In this study, the protein staphylococcal nuclease (SNase) has been used as a model system for the introduction of amino acid substitutions, with the goal of creating an internal metal ion binding site. SNase does not typically bind to a metal ion in its beta barrel interior, the location of the desired switch in this study, and so a comparison between the engineered protein and other variants of SNase may yield information on metal ion binding and protein folding processes. SNase was selected

because it has been an effective vehicle in protein electrostatics studies, with established protocols for genetic modification and product purification, yielding a collection of several hundred variant proteins characterized spectroscopically and structurally [4-6].

Engineering a novel metal ion binding site remains a formidable challenge. Current algorithms are unable to meet the requirements of this challenge due to the number of parameters necessary in protein folding around a metal ion [7]. Ligation of positively charged metal ions requires specific interactions with amino acid side chains that carry a negative charge or a lone pair of electrons. Insertion of these ionizable amino acids and the metal ion into the hydrophobic interior of the protein imposes a severe thermodynamic penalty for the protein, decreasing its structural stability. Hyperstabilized SNase variants, created by incorporation of other, stabilizing amino acid substitutions in the protein, shift the protein's baseline stability higher to accommodate subsequent destabilizing substitutions. Consequently, many SNase protein variants retain a native fold despite containing one or more internal ionizable amino acids [5]. The use of hyperstabilized backgrounds has allowed SNase to function as a model system for protein electrostatic studies. It is small and monomeric, with established protocols for genetic modification and protein purification [4-5]. Consequently, several hundred SNase variants have been characterized spectroscopically by means of fluorescence, circular dichroism, and nuclear magnetic resonance techniques, and structurally by X-ray crystallography. A hyperstabilized SNase variant was selected as the starting protein for this study.

Engineering a metal ion dependent switch into a protein would enhance understanding of protein folding and dynamics. Development of a successful switch protein could provide a framework for studying mechanisms of metalloprotein folding, metal ion specificity, and metal ion sensing.

Methods

Protein Design

The preliminary metal ion-dependent switch proposed in this study was based on the bacterial protein NikR (figure 3), a nickel(II) responsive regulatory protein, which includes both low- and high-affinity nickel(II) binding sites [6]. NikR binds to DNA to modulate the expression levels of genes encoding proteins that protect the organism from potentially toxic nickel (II) ions (Ni^{2+}). Its maximal activity is therefore achieved only in the presence of the metal ion that it is meant to help remove from the cell, by activating transcription of the gene encoding proteins that expel the metal ions. The high-affinity binding site of NikR consists of three histidine residues, each of which contains an ionizable imidazole side chain, and a single cysteine residue, which is a polar amino acid with a reactive thiol group. This binding site is highlighted in the zoomed inset of figure 3, with the Ni^{2+} ion in green ligating to a nitrogen atom of each of the three histidines in blue and the sulfur atom of the cysteine shown in yellow. The site spans two protein subunits, shown in pink and orange in figure 3, and has a square planar geometry. Binding affinity is measured by a K_d value, given in molarity, which represents the equilibrium dissociation constant. Ni^{2+} binding affinity in *Pyrococcus horikoshii* NikR is in the nanomolar range (10^{-9} M), and in *Escherichia coli* NikR, the affinity is in the picomolar range (10^{-12} M) [8-9]. These small values indicate a very high binding affinity for NikR to the Ni^{2+} ion.

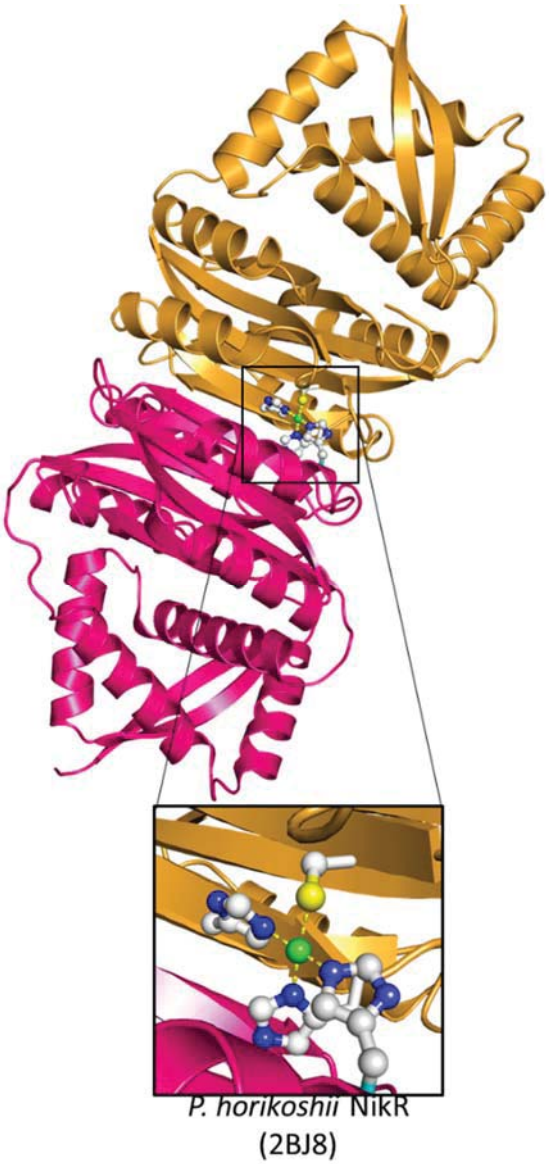

Figure 3: Ribbon diagram of *Pyrococcus horikoshii* NikR (2BJ8). The Ni^{2+} ion is ligated in a square planar geometry by four amino acid side chains shown in the inset. The figure was generated using Pymol [10].

Visual inspection of the crystallographic coordinates of a hyperstabilized variant of SNase (Δ+PHS, Protein Data Bank ID 3BDC) using Pymol software [10] was performed to identify potential sites for the four Ni^{2+} ligands. Superimposed crystallographic coordinates and thermodynamic data for previously characterized SNase variants containing single histidine substitutions were used to guide rotamer preferences for amino acid side chains to optimize the metal ion binding site and minimize steric clashes. A sample calculation to determine thermodynamic stability is provided in equation 1.

$$\Delta G^{\circ}_{H2O,variant} = \Delta G^{\circ}_{H20,background} - \sum \Delta\Delta G_{H2O,variant}$$

Equation 1: Thermodynamic stability of protein variant computed by subtracting the thermodynamic penalty of amino acid substitution from the protein background stability.

Selected amino acids were also selected for truncation to alanine to provide additional space in the protein interior. Amino acid substitutions at the following positions were selected: valine-23, leucine-25, phenylalanine-34, leucine-36, valine-66 and isoleucine-92. Each position had been substituted singly in a previous study in the Garcia-Moreno lab, providing thermodynamic, and in some cases structural, information. Corresponding reference proteins that included the substitutions introduced to enhance thermodynamic stability or create space for the metal ion binding site, but which lacked the metal ion ligands, were also designed. Estimates for the thermodynamic stability of the variants designed in this study were made by adding experimentally determined thermodynamic costs for each individual amino acid substitution (J. Sorenson, Ph.D. dissertation, Johns Hopkins University, 2015). In addition, a more highly stabilized background variant, Δ+NVIAGLA, was selected for the initial design strategy to provide additional stabilization against the substituted ligating residues. Δ+NVIAGLA has been

demonstrated to have an additional 6.7 kcal/mol of thermodynamic stabilization (as Gibbs free energy change of unfolding, $\Delta G°_{H2O}$) relative to Δ+PHS (Δ+NVIAGLA: $\Delta G°_{H2O}$ = 13.2 kcal/mol; Δ+PHS: $\Delta G°_{H2O}$ = 11.8 kcal/mol) [11-12]. A later design utilized the Δ+PHS background. In this convention, a more positive value indicates a more stable protein and a negative value indicates an unstable protein, likely to be unfolded. The first three candidate switch proteins had unfavorable estimated $\Delta G°_{H2O}$ values, predicting that each protein would be unfolded in the absence of metal ion. The final candidate, with a Δ+PHS background, had a near-equilibrium stability estimate. Each reference protein had a favorable estimated $\Delta G°_{H2O}$ value, predicting that the protein would be folded.

DNA Plasmid Production

To introduce amino acid substitutions into the SNase protein, changes were made in the nucleotide sequence of the gene encoding SNase. This gene was located on a pET24a(+) plasmid (Novagen), a double-stranded circular DNA containing a kanamycin resistance gene and T7 promoter. For each round of substitutions, two oligonucleotide primers were designed and produced commercially (Integrated DNA Technologies) to introduce nucleotide changes in the gene sequence that corresponded to the desired amino acid substitutions in Δ+NVIAGLA or Δ+PHS. A polymerase chain reaction (PCR) was performed using these primers, the plasmid DNA template, mixed deoxyribonucleotides (equimolar dATP, dGTP, dCTP, and dTTP), *Pfu* Turbo DNA polymerase (Agilent Technologies), and appropriate buffer. During this PCR experiment, an order of 10^7 copies of the altered plasmid were produced. Following electrophoresis of PCR products on 0.5% agarose gels to confirm success of the plasmid amplification, the PCR reaction mixtures were treated with DpnI enzyme for two hours at 37°C to digest the original DNA template and then transformed into competent *E. coli* BL21 (DE3)

cells. Following successful growth of the transformed cells on growth media containing 30 μg/mL kanamycin to ensure uptake of the modified plasmid, the plasmids were purified from the host cells using a Wizard miniprep kit (Promega) and sequenced commercially (Genewiz) to confirm the substitutions. Iterative rounds of PCR, gel electrophoresis, DpnI digest, transformation, plasmid purification and gene sequencing were performed until each final gene sequence was obtained. Bacterial strains were maintained as glycerol stocks at -80°C.

Protein Expression and Purification

This study used a purification protocol developed in the García-Moreno lab, modified from the method of Shortle and Meeker, to produce and purify the engineered variants of staphylococcal nuclease [4-5]. Proteins were expressed and purified by ENS Nic Butler (USNA '16), 1/C Kilen, and Prof. Schlessman. First, the *E. coli* cell cultures with the modified plasmid coding for the desired SNase variant were grown. Using a glycerol stock of the bacterial strain stored at -80°C, cells were streaked onto a plate containing lysogeny broth (LB), a nutritionally rich environment for cells to grow, and 30 μg/mL kanamycin to ensure that only the desired cells were grown. After an overnight incubation of the plate at 37°C, a single colony was selected to inoculate the liquid culture. The use of a single colony ensured that the bacterial culture was genetically identical. This colony was used to inoculate 1.5 L of sterilized LB containing 30 μg/mL kanamycin. The cells were grown overnight at 37°C with shaking at 220 rpm. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to the flask (final concentration 1mM IPTG). IPTG mimics allactose and so acts to de-repress the bacterial *lac* repressor protein which constitutively inhibits transcription *in vivo* [13], thereby allowing DNA-dependent RNA polymerase native to the *E. coli* host cells to bind to the promoter site on the plasmid and initiate transcription of the target gene. The mRNA produced in transcription

provides the transcript from which the protein was synthesized in the cell by the host cell's ribosome. Cell growth continued for 4 hours after inducing with IPTG. Due to the large quantity of target protein produced by the cells, these proteins were sequestered in inclusion bodies in the cells to minimize impact to the remaining cellular operations.

To extract the target protein from the host cells, a series of centrifugation and incubation steps was performed. First, the cells were collected by centrifugation for 10 min at 5000 rpm and 4°C in a floor centrifuge. The supernatant was discarded, and the cell pellets were resuspended in buffered solution containing 6M urea at pH 8 (extraction buffer 1). Following centrifugation, the supernatant was discarded and the pellet was resuspended in a second buffer containing 6M urea and NaCl at pH 8 (extraction buffer 2). This buffer served to free the target protein from inclusion bodies that were precipitated in the previous step. An additional centrifugation step resulted in the target protein in the supernatant, which was decanted and saved.

The next step of purification involved stepwise precipitation of unwanted protein and nucleic acids, using ice-cold ethanol added to the protein solution followed by incubation at -20°C. Initially, a lower concentration of ethanol was used to precipitate contaminants such as nucleic acids, while the target protein remained in solution [14]. Centrifugation of the mixture was followed by decanting of the supernatant containing the target protein. The supernatant was retained and the pellet was discarded. A second portion of ice-cold ethanol was added to the supernatant and the mixture was incubated at -20°C to precipitate the target protein. The mixture was centrifuged and the supernatant was discarded, leaving a pellet containing the desired protein. This pellet was resuspended in 30 mL of extraction buffer 1. The protein solution was applied to a 5 mL column of SP Sepharose chromatographic resin (G.E. Healthcare) pre-equilibrated with extraction buffer 1, followed by washing of the column with an additional

100mL of extraction buffer 1. Protein was eluted from the column using 40mL of extraction buffer 2. Addition of 160mL of ice-cold ethanol to the eluted protein followed by incubation at -20°C was performed to precipitate the target protein, which was later harvested by centrifugation under the same conditions as the previous step. The pellet was resuspended in extraction buffer 2 and loaded into prepared 8,000-10,000 MWCO dialysis tubing (Spectrum). Dialysis of the resuspended protein in 4L of 1M KCl was performed overnight at 4°C to lower the concentration of urea surrounding the protein and promote protein refolding. Stepwise reduction of the KCl concentration in the dialysis buffer was performed over several days until the protein was dialyzed into two exchanges of pure water.

The target protein was harvested from dialysis by first centrifuging the solution at 4°C to remove any precipitated protein. The supernatant was collected and its concentration was estimated by absorbance at 280nm. Protein was frozen dropwise into liquid nitrogen and stored at -80°C for future use. An aliquot of the protein was used to determine its purity, which was assessed using gel electrophoresis under denaturing conditions. The background and candidate proteins, and molecular weight standards, were loaded onto 15% polyacrylamide gels in the absence and presence of the reducing agent β-mercaptoethanol (βME). An electrical field was applied to separate the proteins based on molecular weight (figure 4). Gels were stained using Coomassie blue and destained prior to viewing.

Additional Protein Refolding Experiments

For the third candidate protein, dialysis with 2.0M guanidinium chloride solution containing 50mM Tris (pH 8), 100mM KCl, and 25mM $NiCl_2$ was performed in an attempt to direct protein refolding. The goal of this experiment was to render a partially-folded protein to a fully unfolded state. Incubating this fully unfolded protein with excess Ni^{2+} ions would provide

the protein an opportunity to bind to the metal ion and adopt its native fold. The concentration of denaturant was then lowered gradually until completely removed.

Refolding was also attempted through flash dialysis. To prevent proteins from aggregating, 3 mL of candidate protein was added dropwise to 50 mL of buffer solution containing 50mM Tris (pH 8), 100mM KCl, and 25mM $NiCl_2$ at a rate of 100 µL of protein per minute. The resulting solution was concentrated in a floor centrifuge using 15 mL Amicon Ultra-15 centrifugal concentrators of MWCO 10,000 (Millipore). Protein concentration for the final solution by each method was measured using absorbance at 280nm.

The effects of pH were tested over an array of pH from 7-9 using Tris buffer titrated to each pH value. Ionic strength was maintained using 0.10M KCl. These experiments included the switch candidates and their reference proteins, as well as SNase variants known to adopt a native fold, Δ+PHS, Δ+PHS/V23H, and Δ+PHS/L36H, and a SNase variant known to exist in a fully-unfolded state, WT/T62P. These protein variants served as positive and negative controls, respectively. These proteins were incubated separately in the absence and presence of 25mM $NiCl_2$ prior to spectroscopic analysis.

Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy was used to detect conformational changes between folded and unfolded SNase proteins by examining secondary structural features. At USNA, a Jasco J-815 CD spectrometer was used. Additional spectra were collected at JHU using an Aviv model 215 CD spectrometer. Spectra were collected over the wavelength range from 260 to 200 nm. This technique uses circularly-polarized UV light that is transmitted differently by chiral molecules such as proteins [15]. Proteins with different elements of secondary structure, such as alpha helices or beta sheets, have characteristic CD spectra that are

different from those of unfolded, “random coil” proteins. Spectra for each secondary structure element result from $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ electronic transitions [15]. SNase is predominantly composed of three alpha helices in addition to a five-stranded beta barrel core. However, the alpha helical signature is stronger and dominates the signal. Spectra collected over the wavelength range from 260 to 200 nm using native-folded SNase proteins revealed a characteristic “W” shape with negative peaks at 222 and 208 nm (figure 4). These characteristics provide a signature of a protein containing alpha helices. Proteins in which secondary structure has been lost will lack these spectral features and, instead, their spectra will contain a broad negative signal at wavelength centered at 195nm. Background spectra containing only the buffer were also collected for each set of conditions. Protein concentration was measured for each CD sample using absorbance at 280nm. In Excel, the background signals were subtracted from signals for each protein and the resultant signals were normalized for protein concentration and molar mass to yield a plot of mean residual ellipticity as a function of wavelength.

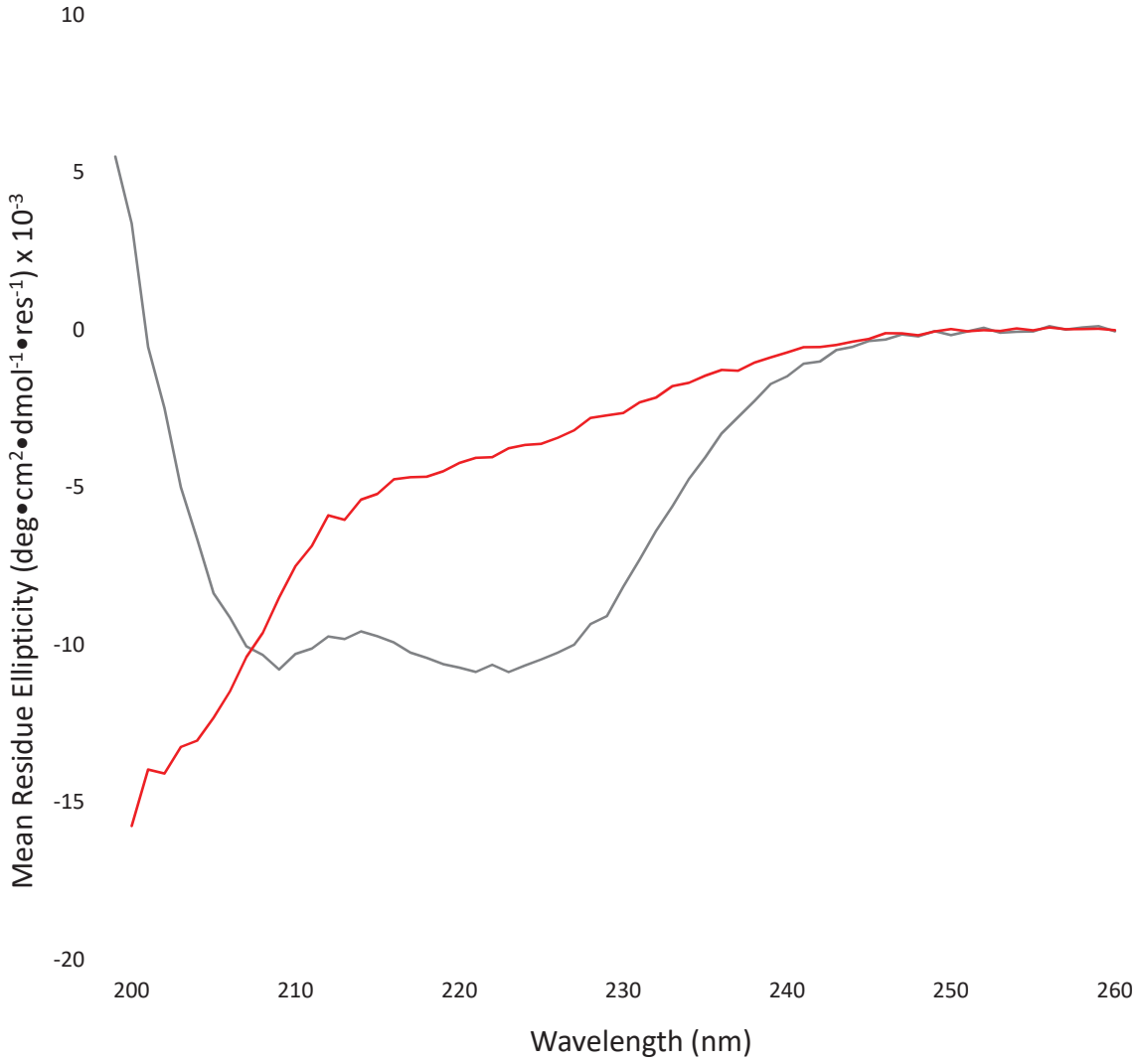

Figure 4: CD spectra depicting folded and fully unfolded SNase protein secondary structure signatures. The "W" shape of Δ+PHS (grey) displays an α-helical signature and the shape of the WT/T62P variant (red) depicts complete random coil.

X-ray Crystallography

X-ray crystallography provides near-atomic resolution structural information for macromolecules such as proteins. To gain skill in growing protein crystals, performing diffraction data collection and processing, and crystallographic computational methods, a structure determination was undertaken using the SNase variant Δ+PHS/V66Y. This protein was supplied by the García-Moreno group. Hanging drop vapor diffusion methods were used to identify the initial crystallization conditions, beginning from conditions known to crystallize other SNase variants, and then optimized to produce crystals suitable for high-resolution diffraction studies. Initial reservoir solutions included 24 conditions of 20-45% 2-methyl-2,4-pentanediol (MPD), in 5% increments, and 25mM KH_2PO_4/K_2HPO_4, pH 6-9, in 1 pH unit increments. Refined conditions utilized 1% MPD increments. To set up each crystallization experiment, the purified protein, initially at 13mg/mL concentration, was combined with $CaCl_2$ and the inhibitor thymidine-3',5'-diphosphate (THP) to yield a molar equivalents of 1 protein : 3 $CaCl_2$: 2 THP, prior to mixing 4μL of the protein solution with 4μL of the reservoir solution on a siliconized cover slip. The cover slip was inverted over a well containing 1mL of the reservoir solution and sealed with vacuum grease. Crystallization experiments were incubated at 4°C and inspected weekly for crystal growth. Candidate crystals were suspended in rayon loops attached to magnetic bases, flash-cooled in liquid nitrogen, and stored submerged in liquid nitrogen until they were screened by the Bruker DUO diffractometer system in the USNA Chemistry department. Diffraction data were collected for two different crystals. The data were processed using manufacturer's software to yield a set of unique reflections. Molecular replacement methods using published atomic coordinates for a related SNase protein (PDB ID: 3BDC) were used to provide initial phasing for the data set [10]. Iterative model building and crystallographic

refinement were performed to yield atomic coordinates, which included residues 7-141 of the protein, one Ca^{2+} ion, 1 THP ligand, and water molecules [16-19]. The final coordinates will be published in the Protein Data Bank (www.pdb.org) and analyzed in the García-Moreno group by comparison to related SNase protein structures and will be included in a peer-reviewed article. An analogous experimental plan would be used for characterizing successful metal ion-binding switch candidate proteins and reference proteins.

Results

To date four SNase variants containing a metal ion binding site based on the NikR high-affinity binding site have been designed. Three variants were synthesized into a previously hyperstabilized background protein, Δ+NVIAGLA. Another variant possessed two substitutions in the Δ+PHS SNase background. Mutations of the gene encoding SNase have been introduced to encode these substitutions. The mutated SNase protein and its corresponding background protein have been expressed in *E. coli* cells and purified to allow for characterization. The assessment of protein purity is given in figure 5.

The gel provided assessed protein purity by comparing the expressed protein to molecular weight standards. Proteins of different weights are separated by electrophoresis. Lower bands indicate lower molecular weights. SNase in Δ+PHS and Δ+NVIAGLA variants migrate to 16 kiloDaltons. Wild type SNase has a greater molecular weight and so produces a band of strong intensity higher than 16 kDa.

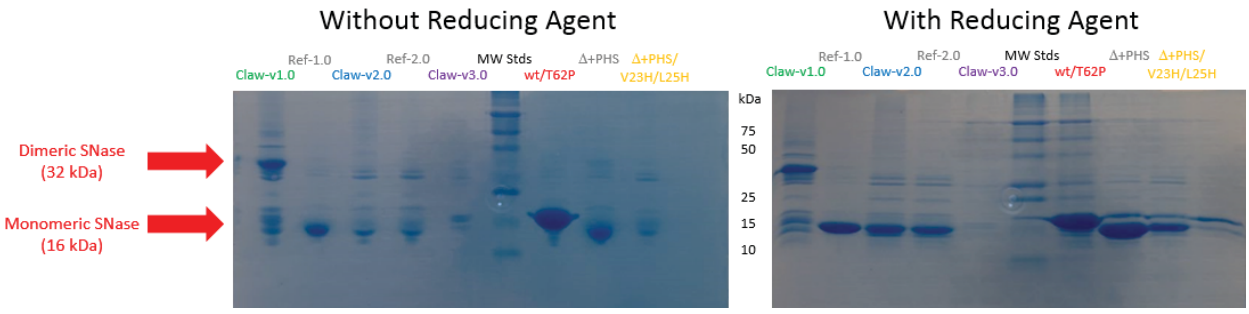

Figure 5: Denaturing electrophoretic gels of SNase variant proteins run in the absence (left) and presence (right) of the reducing agent, β-mercaptoethanol. Bands at 16 kDa indicated a monomeric protein. Bands at both 16 kDa and 32 kDa in the left panel suggested a disulfide-linked dimer formed between the single cysteine thiol from adjacent monomers in the Claw-v1.0 and Claw-v2.0 variants. Background proteins, Δ+PHS, WT/T62P, Claw-v3.0 and V23H/L36H did not contain cysteine residues. Molecular weight standard masses are indicated.

Claw-v1.0

Five amino acid substitutions were introduced into the β-barrel core in the SNase protein interior, as shown in figure 6 [8]. The metal ion binding site was based on the square-planar high-affinity Ni^{2+} site found in NikR [6] and included one cysteine and three histidines. Crystallographic coordinates for three single-substituted SNase variants, with a histidine introduced at position 25, 36, or 66, were superimposed onto the Δ+PHS coordinates in Pymol [10], revealing a first approximation for the three histidines in the NikR high-affinity Ni^{2+} site. The PDB ID codes for these coordinates are 5C4Z (Δ+PHS/L25H), 4LAA (Δ+PHS/L36H), and 5C3W (Δ+PHS/V66H). Position 34 in SNase was selected for the Cys ligand and was introduced into the model coordinates using a mutagenesis tool in Pymol [10]. Complex amino acid side chains may adopt several positions, called rotamers. The cysteine rotamer was selected from a library of three options to optimize the geometry of the Ni^{2+} binding site. To minimize steric clashes in the putative Ni^{2+} binding side, a valine at position 23 was substituted with a smaller alanine residue. This substitution was included in both the candidate switch protein as well as in the reference protein used for comparison. The final protein Δ+NVIAGLA/V23A/L25H/F34C/L36H/V66H) was the design for claw-v1.0. The choice of these positions is significant as rotamers for the three histidine residues added have been determined using X-ray crystallography, so the position of these residues was known experimentally. An amino acid sequence alignment between wild type (naturally-occurring) SNase and the proteins used in the study is shown in figure 7.

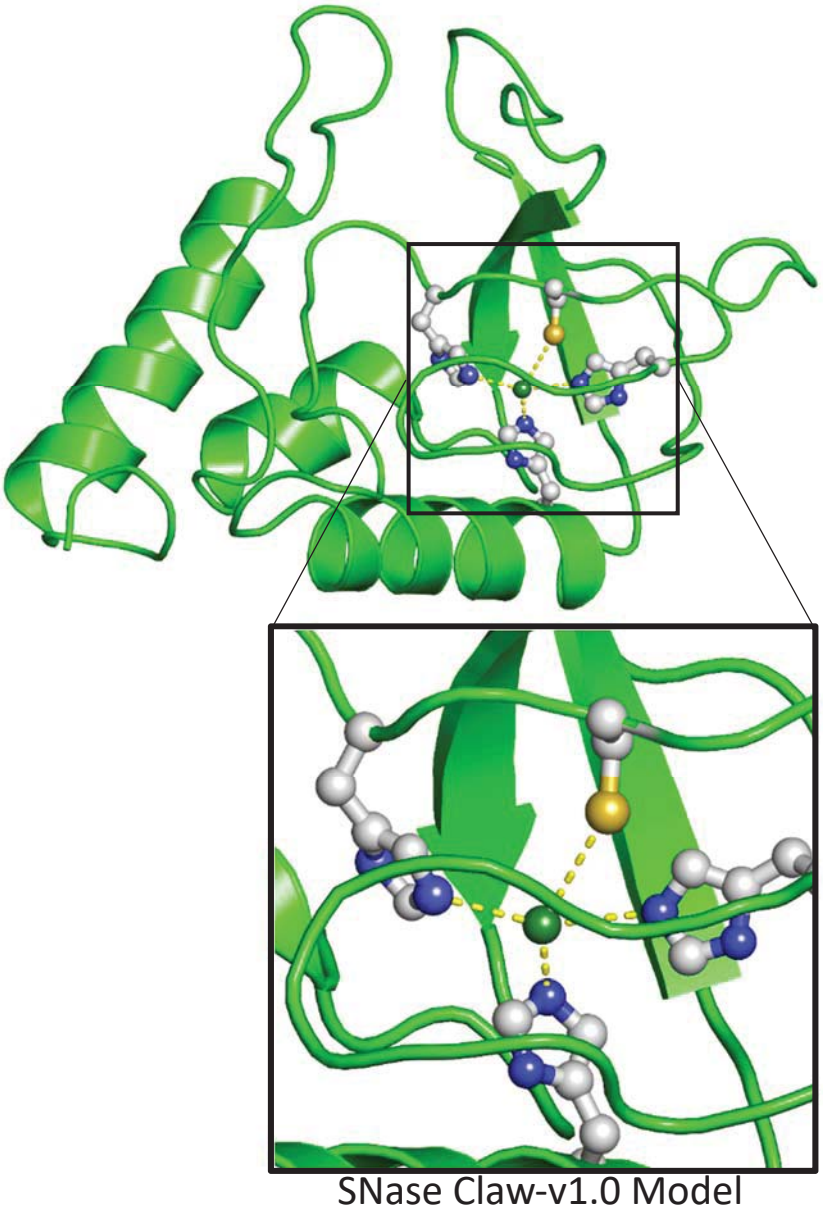

Figure 6: Ribbon diagram of a hypothetical model of claw-v1.0. The Ni^{2+} binding site is positioned in the SNase β-barrel. The sulfur atom in the cysteine is shown in yellow and the nitrogen atoms in each histidine are shown in blue. The figure was generated using PyMol [10].

Estimates of the Gibbs free-energy changes of each protein relative to wild-type SNase have been calculated by adding empirically-determined thermodynamic penalties for individual amino acid substitutions. The reference protein, Δ+NVIAGLA/V23A, had an estimated Gibbs free-energy change of unfolding, $\Delta G°_{H2O}$, 10.5 kcal/mol, indicating a strong likelihood that the protein would be folded under experimental conditions. In contrast, the candidate protein claw-v1.0 with the designed metal switch had an estimated $\Delta G°_{H2O}$ of -10.2 kcal/mol, suggesting that it would not retain the native fold of SNase.

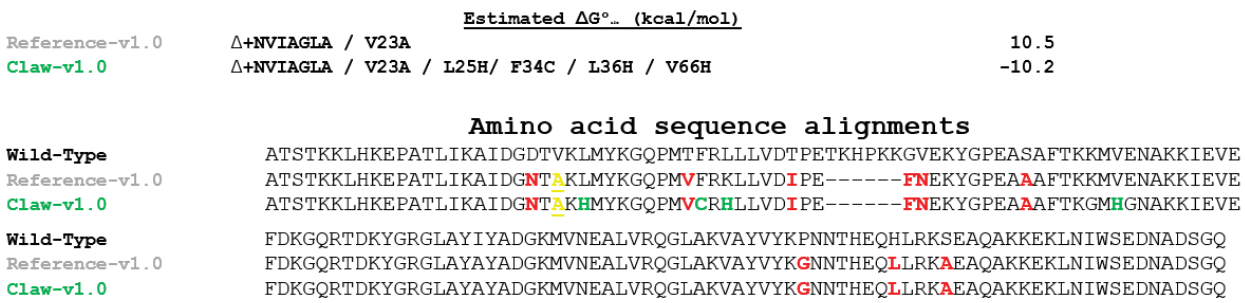

Figure 7: Amino acid sequence alignment for the background protein and claw-v1.0 with wild-type (naturally-occurring) SNase. Each amino acid is represented by its one-letter abbreviation (ex: A = alanine). Amino acids that have been altered from the wild-type sequence are in red. Sequences changed from the Δ+NVIAGLA background are shown in yellow. Amino acids that have been changed relative to the background sequence to make the claw-v1.0 protein are depicted in green.

By assessing the protein purity with denaturing gel electrophoresis in the absence and presence of the reducing agent β-mercaptoethanol, it was determined that the protein was likely forming a disulfide bridge with a neighboring protein molecule. Disulfide bridges occur between adjacent thiol groups, such as cysteine side chains in proteins. In the left panel of figure 8, protein samples underwent denaturing gel electrophoresis in the absence of a reducing agent. In this experiment, the claw-v1.0 migrated as two bands at 16 kilodaltons (kDa) and 32 kilodaltons (kDa), where kilodaltons are a unit of molecular weight given in kilograms/mole. A band at 32

kDa indicates that the SNase may have formed a disulfide bridge, twice the expected weight of a monomeric SNase protein. The background protein, in contrast, migrated as a single band at 16 kDa, consistent with the known molecular weight of one SNase protein subunit. The amino acid cysteine, which contains a reactive thiol group, is not found in the wild-type and background SNase amino acid sequence. The introduction of a single cysteine into the unfolded metal ion-dependent switch protein candidate may have resulted in disulfide bond formation between two unfolded monomers under oxidizing conditions, yielding a protein dimer of molecular weight 32 kDa. The disulfide bond would have persisted during dialysis after the original protein purification. The right panel of figure 8 shows a denaturing gel run in the presence of β-mercaptoethanol, which reduces disulfide bonds. Here, both the background and claw-v1.0 proteins migrated as a single band at 16 kDa, consistent with monomeric SNase. These results suggested that future spectroscopic experiments should use reducing conditions to avoid interference of the disulfide bond with Ni^{2+} binding and protein folding.

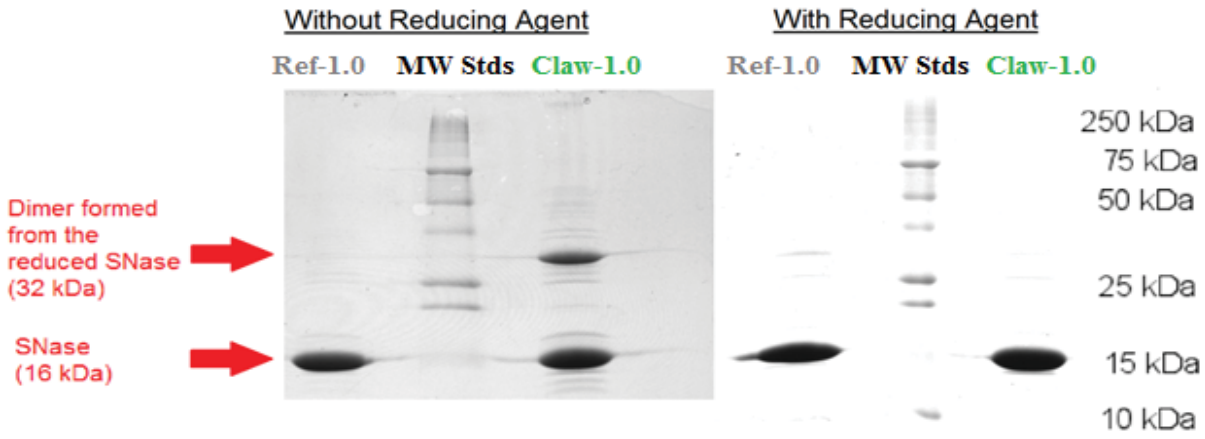

Figure 8: Gel electrophoresis run on both background and claw-v1.0 proteins in the absence (left) and presence (right) of the reducing agent β-mercaptoethanol. The proteins were expected to migrate through an electric field as a monomer of 16 kDa. However, the claw-v1.0 protein migrated as two bands as 16 kDa and 32 kDa, suggesting that a portion of that protein existed in a dimeric form. Molecular weight standard masses are indicated.

Circular dichroism spectroscopy was performed to obtain information about the claw-v1.0 and reference protein secondary structure under varying conditions. The initial spectra showed that the background protein retained a native-fold while claw-v1.0 was partially unfolded (figure 9.A). The background protein spectra (grey) was consistent with the signature of an alpha helical protein, with negative peaks at 222 nm and 208 nm. Reference-1.0 retained this shape after being incubated with Ni^{2+} overnight. Reference-1.0 protein and claw-v1.0 protein spectra had a negative peak at 222 nm, but claw-v1.0 (in green) exhibited a shift to lower wavelengths, indicative of loss of alpha signature. The claw-v1.0 protein was partially folded, as exhibited by the increase in mean residue ellipticity at 204 nm. Compared to the WT/T62P spectra (red), which exhibited the behavior of a completely unfolded protein, the claw-v1.0 protein displayed a partially-folded structure.

A second experiment involved further unfolding of claw-v1.0 using a chemical denaturant (figure 9.B). Claw-v1.0 protein incubated in 0M to 2.0M guanidinium chloride (GdmCl), in 0.5M increments, displayed CD signal loss at 222 nm at even the lowest concentration of the denaturant, 0.5M GdmCl. This loss of signal implies that in the absence of denaturant, the claw-v1.0 protein was partially unfolded, but the residual structure was readily lost in the presence of low concentrations of chemical denaturant. Incubation of the background and claw proteins with molar excess of Ni^{2+} was performed in the presence and absence of the reducing agent β-mercaptoethanol. No change in signal was found for either protein (figure 9.C). The CD signature of the reference protein under conditions of 1000-fold stoichiometric excess of Ni^{2+} and 2mM β-mercaptoethanol (yellow) remain unchanged relative to that of the reference protein in buffer (green). Similarly, the claw protein did not recover a native-like fold upon incubation with Ni^{2+} in the absence or presence of reducing agent. Under these conditions, either

Ni^{2+} binding to the protein did not occur or the binding energy was insufficient to allow the protein to fold. Disulfide-mediated dimer formation in the claw-v1.0 could no longer be considered a limiting factor in refolding when the disulfide bond is reduced by β-mercaptoethanol, as reducing the cysteine residue would prevent the formation of the disulfide bond. However, due to the proximity of the standard reduction potentials for Ni^{2+} (to elemental Ni) and β-mercaptoethanol, a possibility that the reducing agent acted on the metal ion rather than the protein cannot be excluded. Molar excess of β-mercaptoethanol over the protein concentration suggests that at least some of the protein disulfide bonds were likely to have been reduced in the experiment.

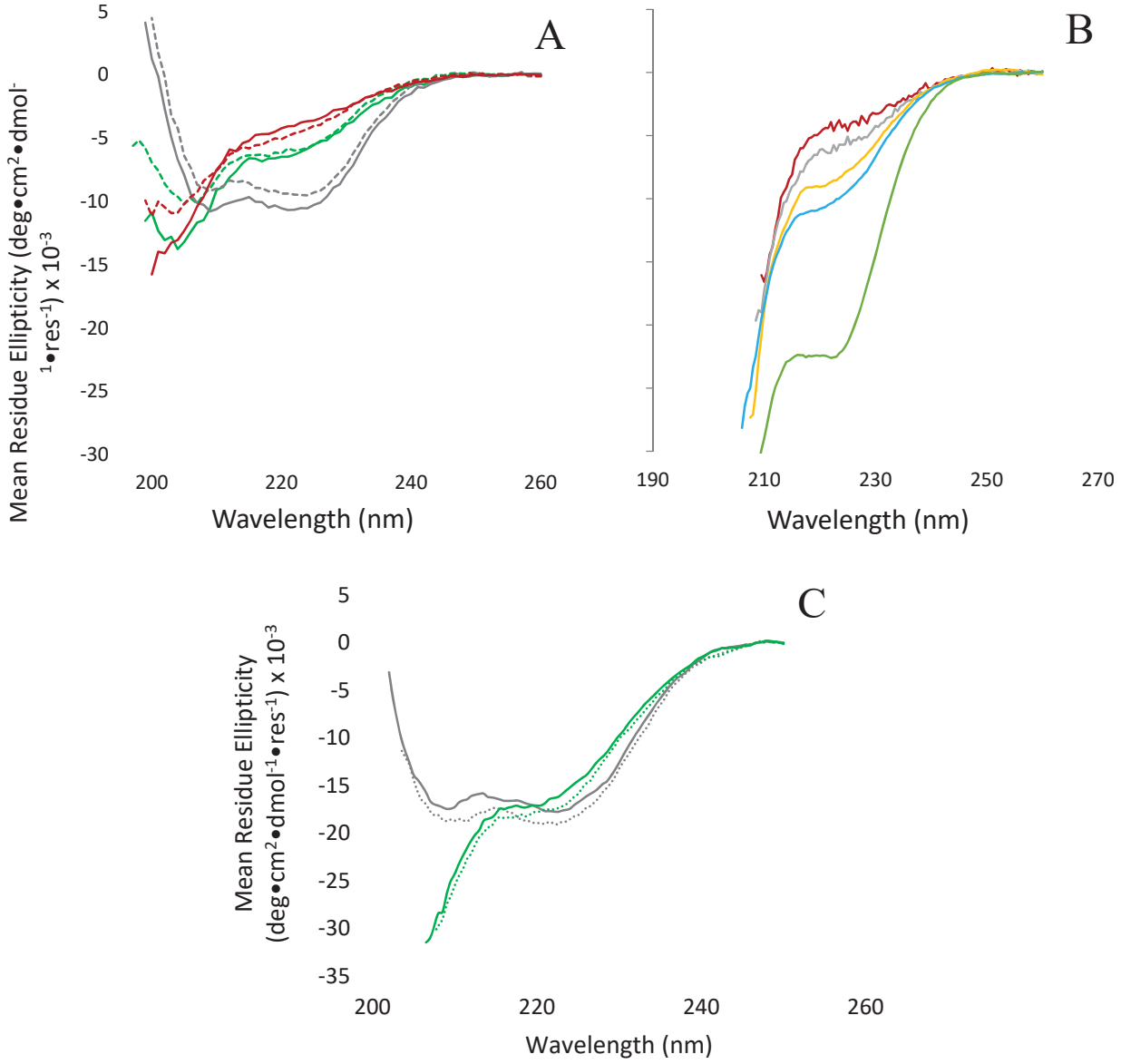

Figure 9.A: Initial CD spectra showing that in the absence of Ni^{2+} claw-v1.0 (green, dotted) is partially unfolded relative to reference-v1.0 (grey, dotted) and WT/T62P (red, dotted) at pH8. Incubation of protein in the presence of Ni^{2+} shows that claw-v1.0 (green, solid), reference-v1.0 (grey, solid) and WT/T62P (red, solid) are unaffected by metal ion. **9.B:** CD spectra showing that guanidinium chloride (GdnCl) denatures claw-v1.0 and to what extent claw-v1.0 protein unfolds as a function of the concentration of chemical denaturant. Without denaturant (0.0M GdnCl, green) the protein's CD signature indicates that it was partially folded. Upon addition of low concentrations of denaturant (0.5M GdnCl, blue) the protein loses a majority of this partial fold. As concentration increases, more of the protein structure is lost (1.0M GdnCl, orange; 1.5M GdnCl, purple; 2.0M GdnCl, red). **9.C:** β-mercaptoethanol does not promote claw-v1.0 refolding in the presence of $Ni^{2=}$ as exhibited by CD spectra of background protein (grey, solid, without reducing agent; grey, dotted, with reducing agent) and claw-v1.0 (green, solid, without reducing agent; green, dotted, with reducing agent) incubated with Ni^{2+}.

Claw-v2.0

After the first iteration of claw design, two subsequent protein variants were designed and produced. Inspection of the initial model of claw-v1.0 revealed that isoleucine at position 92 approached the intended Ni^{2+} binding site within 0.6nm (figure 11.A). This distance may be sufficient to impede the binding of a metal ion to the protein at the target site. Substitution of alanine for isoleucine at position 92 provides additional space for Ni^{2+} to fit in its designed binding site. The reference protein (reference-v1.0, Δ+NVIAGLA/V23A) and claw-v1.0 model coordinates were both modified using Pymol software [10] to incorporate this change. The new proteins were named reference-v2.0 (Δ+NVIAGLA/V23A/I92A) and claw-v2.0 (Δ+NVIAGLA/V23A/L25H/F34C/ L36H/V66H/I92A, figure 11.B).

The alignment of the amino acid sequence for each protein used in this study relative to wild-type SNase is provided in figure 10. The estimated free-energy change of the background protein, 4.0 kcal/mol (reference-v2.0) and 10.5 kcal/mol (reference-v1.0), indicates a strong likelihood that the protein would adopt a native-like fold. In contrast, the claw-v1.0 and claw-v2.0 proteins have negative estimated Gibbs free energy changes, and so would not be expected to retain native-like folds.

		Estimated ΔG°… (kcal/mol)
Reference-v2.0	Δ+NVIAGLA / V23A / I92A	4.0
Claw-v2.0	Δ+NVIAGLA / V23A / L25H / L36H / V66H / I92A	-16.7

Amino acid sequence alignments

```
Wild-Type      ATSTKKLHKEPATLIKAIDGDTVKLMYKGQPMTFRLLLVDTPETKHPKKGVEKYGPEASAFTKKMVENAKKIEVE
Reference-v2.0 ATSTKKLHKEPATLIKAIDGNTAKLMYKGQPMVFRKLLVDIPE------FNEKYGPEAAAFTKKMVENAKKIEVE
Claw-v2.0      ATSTKKLHKEPATLIKAIDGNTAKHMYKGQPMVCRHLLVDIPE------FNEKYGPEAAAFTKGMHGNAKKIEVE

Wild-Type      FDKGQRTDKYGRGLAYIYADGKMVNEALVRQGLAKVAYVYKPNNTHEQHLRKSEAQAKKEKLNIWSEDNADSGQ
Reference-v2.0 FDKGQRTDKYGRGLAYAYADGKMVNEALVRQGLAKVAYVYKGNNTHEQLLRKAEAQAKKEKLNIWSEDNADSGQ
Claw-v2.0      FDKGQRTDKYGRGLAYAYADGKMVNEALVRQGLAKVAYVYKGNNTHEQLLRKAEAQAKKEKLNIWSEDNADSGQ
```

Figure 10: Amino acid sequence alignment for reference 2.0 and claw-v2.0 with wild-type SNase. Amino acids that have been altered from the wild-type sequence are in red. Sequences changed from the Δ+NVIAGLA background are in yellow. Those that have been altered from the background sequence to make the claw-v2.0 protein are in blue.

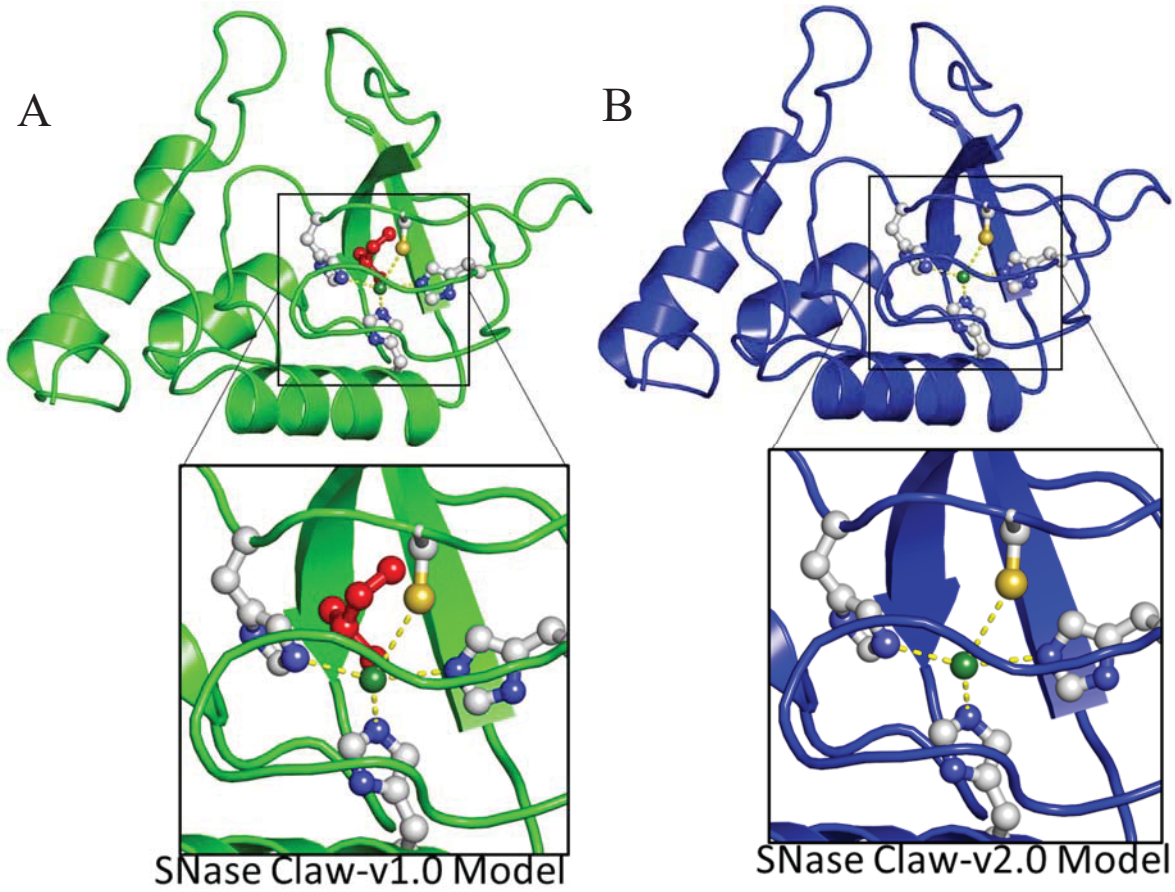

Figure 11.A*:* Ribbon diagram of a claw-v1.0 model with the isoleucine residue depicted at position 92 in red. This residue may interfere with Ni^{2+} binding by extending into the putative binding site. Substitution of isoleucine by alanine may provide additional room for the Ni^{2+} to bind. Therefore, claw-v2.0 replaced isoleucine at position 92 with an alanine. The figure was generated using Pymol [10]. **11.B:** Ribbon diagram of a claw-v2.0 model with the isoleucine residue removed. The figure was generated using Pymol [10].

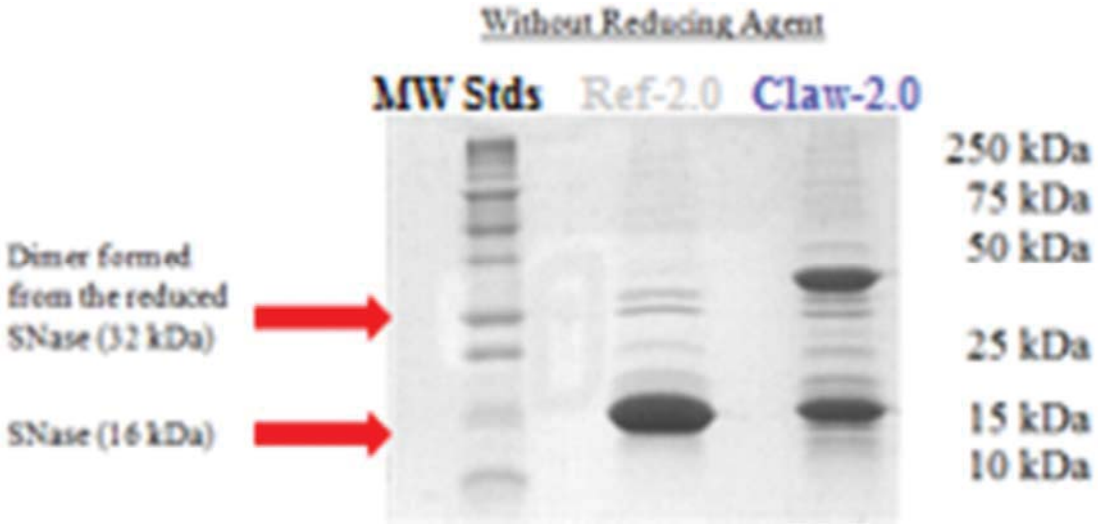

Figure 12: Gel electrophoresis run on both refrerence-v2.0 and claw-v2.0 protein in the absence of reducing agent.

Figure 12 shows the purity assessment of the prepared reference-v2.0 and claw-v2.0 proteins by gel electrophoresis. Here, both the reference-v.20 and claw-v2.0 proteins migrated as a single band at 16 kDa, consistent with the known molecular weight of SNase. Claw-v2.0 also migrated a band at 32 kDa, indicating that a disulfide bridge likely formed under these conditions.

Circular dichroism spectroscopy was performed to assess the secondary structure of claw-v2.0. The reference-v2.0 protein retained a native-fold whereas claw-v2.0 was partially unfolded (figure 13). The spectrum is similar to that in figure 9.A. Incubation of Ni^{2+} with the background and claw-v2.0 proteins also failed to recover the native fold.

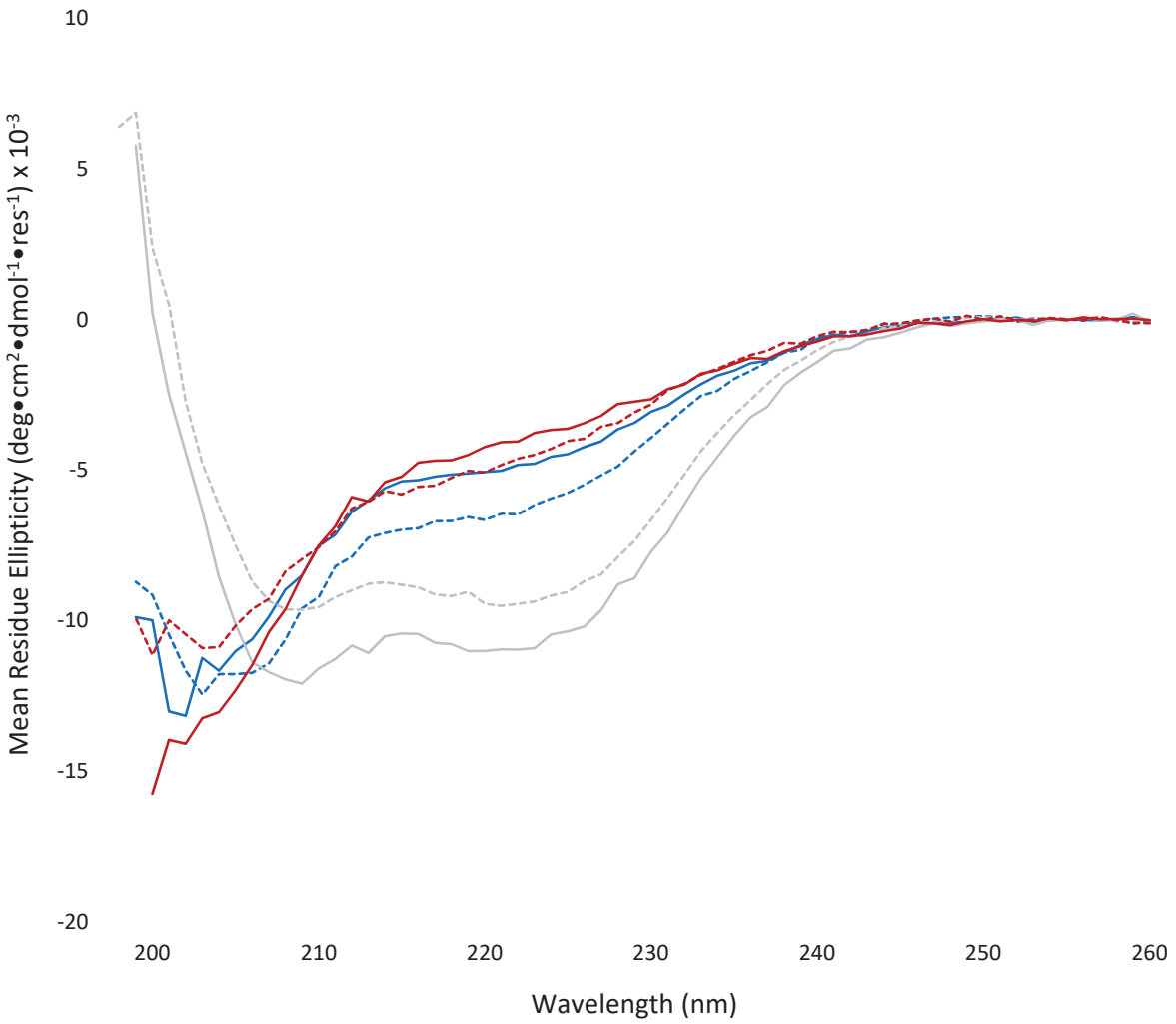

Figure 13: CD spectra of reference-v2.0 (grey, solid), WT/T62P (red, solid), and claw-v2.0 (blue, solid) shows incubation with Ni^{2+} has no effect on protein structure when compared to reference-v2.0 (grey, dotted), WT/T62P (red, dotted), and claw-v2.0 (blue, dotted) CD signature in the absence of Ni^{2+}. Claw-v2.0 retained a partial fold when compared to WT/T62P (red). Studies conducted at pH8.

Claw-3.0

Due to the previous variants' apparent formation of an intermolecular disulfide bond, another candidate protein was designed in an attempt to restore the cysteine at position 34 to phenylalanine (its wild-type sequence). Reverting position 34 to phenylalanine also provided an additional packing surface to mitigate losses from the shortening of the side chain at position 92. The sequence NVIAGLA/V23A/L25H/L36H/V66H/ I92A was designated claw-v3.0. Model coordinates are shown in figure 15. The reference for this protein is background-2.0 (Δ+NVIAGLA/V23A/I92A).

The alignment of the amino acid sequence for each protein used in this study relative to wild-type SNase is provided in figure 14. Note that this iteration is predicted to be more stable than claw-v2.0 (-11.3 rather than -16.7 kcal/mol). This makes sense in light of the fact that one of the substitutions used in previous iterations of protein (F34C) was reverted to its original phenylalanine sidechain. The principle of minimizing substitutions to maintain a higher thermodynamic stability is worth further consideration.

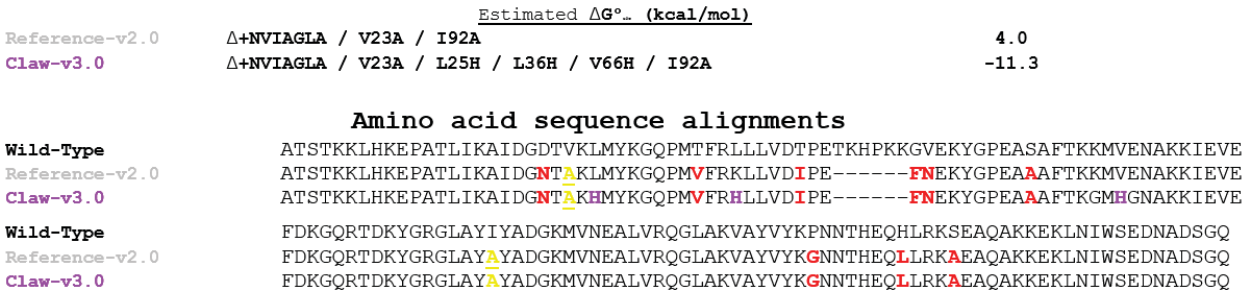

Figure 14: Claw-v3.0 and reference-2.0 amino acid sequence alignment with wild-type SNase. Amino acids that have been altered from the wild-type sequence are in red. Sequences changed from the Δ+NVIAGLA background are in yellow. Those that have been altered from the background sequence to make the claw-v3.0 protein are in purple.

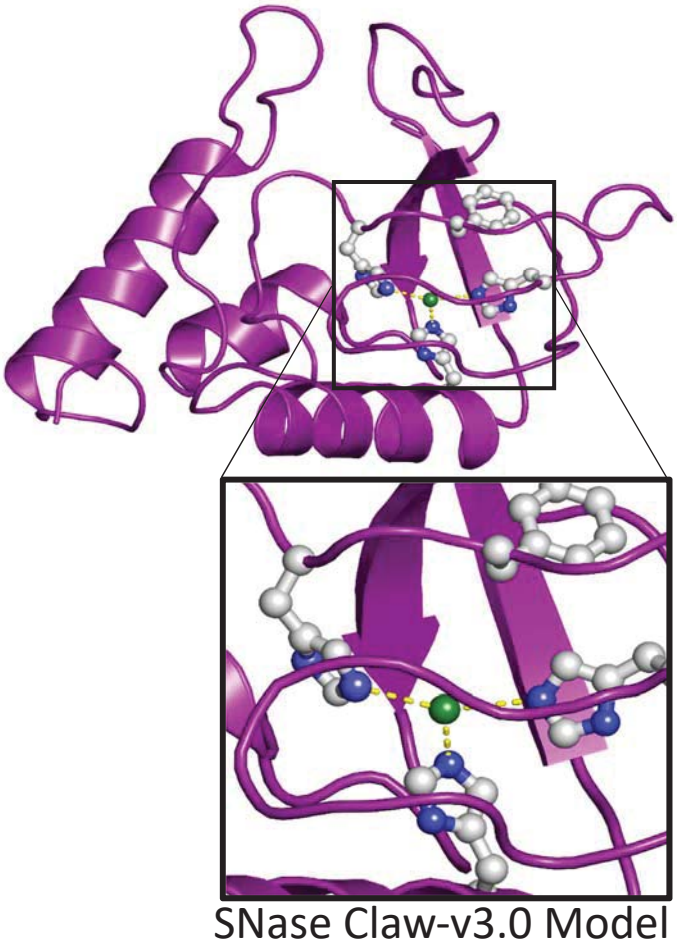

Figure 15: Claw-v3.0 with the cysteine residue restored to phenylalanine at position 34 and with a 180° rotation of the imidazole ring at position 36 to optimize the Ni^{2+} ligation. A putative Ni^{2+} ion is shown in green. The figure was generated using Pymol [10].

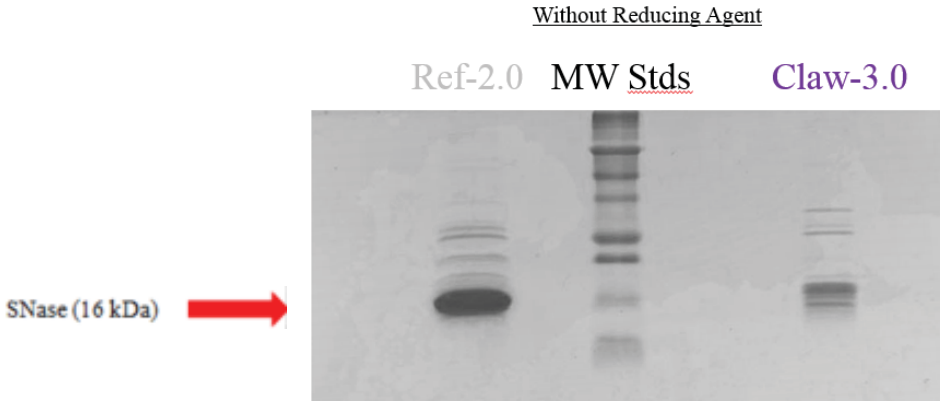

Figure 16: Gel electrophoresis run on both refrerence-v2.0 and claw-v3.0 protein in the absence of reducing agent. There is no band at 32 kDa, consistent with the absence of cysteine in the protein.

Expressing this protein presented significant challenges. The preparation produced a low protein yield. The first three protein preparations were unsuccessful. In the second run, the protein was not produced cleanly, indicated by the smears in the protein gel run from these results. The third preparation yielded a negligible quantity of claw-v3.0. Finally, the fourth protein preparation yielded ~47 mg of claw-v3.0 which was sufficient to collect CD spectroscopic data. Assessment of the protein by gel electrophoresis (figure 16) revealed a single band at 16 kDa and no band at 32 kDa under non-reducing conditions. This migration pattern is consistent with the absence of cysteine in claw-v3.0.

Circular dichroism spectroscopy was used to assess the secondary structure of the third iteration of candidate protein. The reference-v2.0 protein was known to have a native-like fold based on its previous use with claw-v2.0. Claw-v3.0 appeared to be partially unfolded in comparison to the folded reference protein (figure 17.A). The spectra appeared similar to those in figures 9.A and 13 for previous candidate switch proteins. Incubation of the claw-v3.0 proteins in the presence of a 1000-fold molar excess of Ni^{2+} failed to recover a native-like fold. These results suggested a possibility that the protein may already have adopted a stable misfolded conformation in a local energy minimum, so that the thermodynamic requirements to resume a native state were not obtainable, or that the protein was aggregating, thereby preventing binding by the metal ion and protein refolding. Two further studies were designed to examine these effects. The first, a step-dialysis using a chemical denaturant, buffer, and excess Ni^{2+}, would allow the protein to unfold completely in denaturant while being exposed to nickel(II) ions. By gradually lowering the concentration of denaturant, the protein should have been able to bind to the Ni^{2+} and refold. The second method, a rapid dilution, would allow the proteins to be exposed to a large number of Ni^{2+} while keeping the protein concentration low enough that the protein

could not aggregate.

Should the protein have adopted a thermodynamically stable misfolded state, simple incubation with nickel ion would not have allowed it to regain its native fold. To remove this possible intermediate, claw-v3.0 underwent dialysis. Protein harvested from the step dialysis experiment showed no precipitation. However, this experiment failed to restore claw-v3.0 to its native fold, as seen by CD spectroscopic analysis in figure 17.B. Flash dilution was also performed to confirm that protein aggregation was not preventing Ni^{2+} from binding to claw-v3.0. Failure to recover a native fold suggests that the candidate switch proteins may have lower binding affinities to Ni^{2+} than predicted. Incubation of the proteins using metal ions with larger ionic radii might be useful in this instance, as the estimated ligand-to-Ni^{2+} coordination distances were ~0.1nm longer than those in the NikR crystallographic coordinates. Longer distances between ligands and the metal ion would weaken the interactions. Another possibility is that the proteins were too thermodynamically unstable to allow protein:Ni^{2+} binding energies to drive refolding. In this case, generation of a successful metal switch would require fewer destabilizing amino acid substitutions.

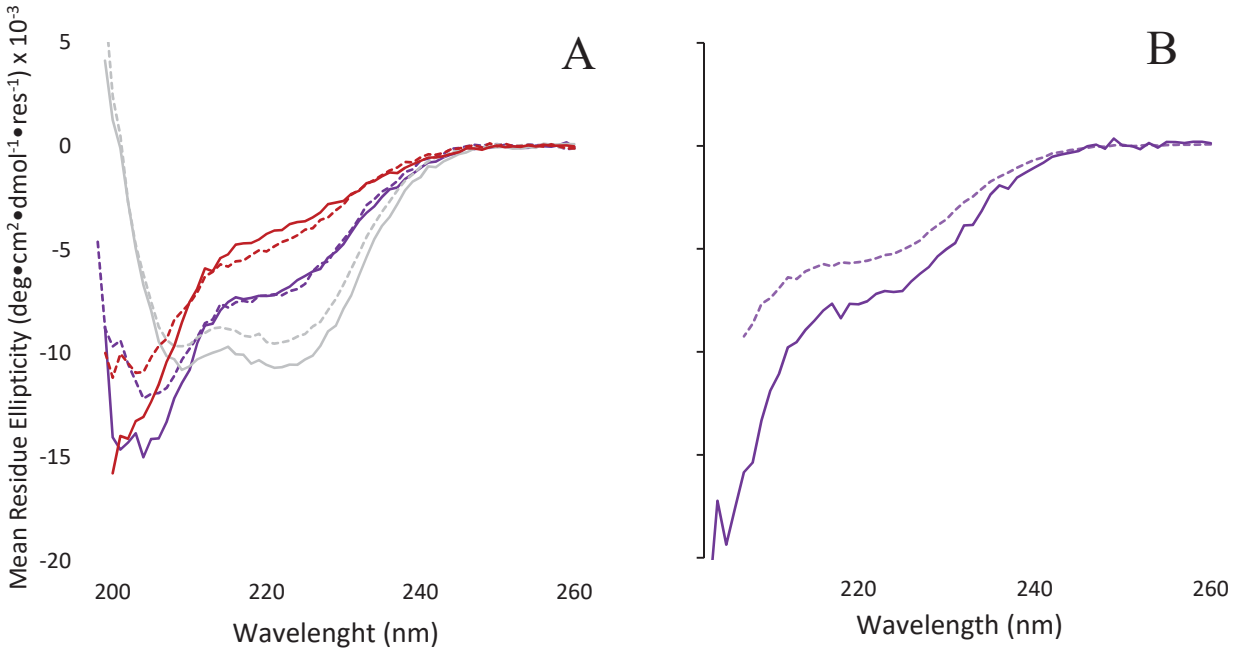

Figure 17.A: CD spectra showing that in the absence of Ni^{2+} claw-v3.0 (purple, dotted) is partially unfolded relative to reference-v2.0 (grey, dotted) and WT/T62P (red, dotted) at pH8. The spectra indicate that incubation in the presence of Ni^{2+} has no effect on claw-v3.0 (purple, solid), reference-v2.0 (grey, solid), or WT/T62P (red, solid). 17.**B:** CD spectra of claw-v3.0 (purple, dotted) indicates that the protein did not recover a native-like fold after being subjected to dialysis in a denaturant and Ni^{2+} and subsequent removal of the guanidinium chloride denaturant. This removes possibility of folding intermediate. CD spectra of claw-v3.0 (purple, solid) diluted in Ni^{2+} then concentrated also failed to recover native fold, eliminating possibility of protein aggregation.

Δ+PHS/V23H/L36H

In an effort to determine the point at which the thermodynamic penalty was so great that it prevented the protein from recovering a native fold, a plan to increase incrementally the number of amino acid substitutions in the protein was developed. SNase variants containing single histidine substitutions into the hyperstabilized Δ+PHS background had been characterized previously (J. Sorenson, Ph.D. dissertation, Johns Hopkins University, 2015). The variant Δ+PHS/V23H/L36H was designed using superimposed crystallographic coordinates for Δ+PHS (PDB ID 3BDC), Δ+PHS/V23H (PDB ID 4ZUI) and Δ+PHS/L36H (PDB ID 4LAA). Beginning from the crystallographic coordinates for Δ+PHS, the valine at position 23 and the leucine at position 36 were both changed *in silico* to histidine. The coordinates shown in figure 18 depict that there are two possible directions the histidine could face at position 23, as published for the Δ+PHS/V23H crystal structure. Choosing rotamer A, which faced the intended Ni^{2+} binding site, resulted in a reasonable coordination geometry between the protein and Ni^{2+} was produced with modest adjustments to the established crystallographic coordinates (4ZUI and 4LAA). The predicted coordination distance for ligation between 23H is 2.1 Å and the coordination distance for ligation with 36H is 2.0 Å. These distances are within the coordination distances observed for the high-affinity Ni^{2+} site in NikR crystal structures [6].

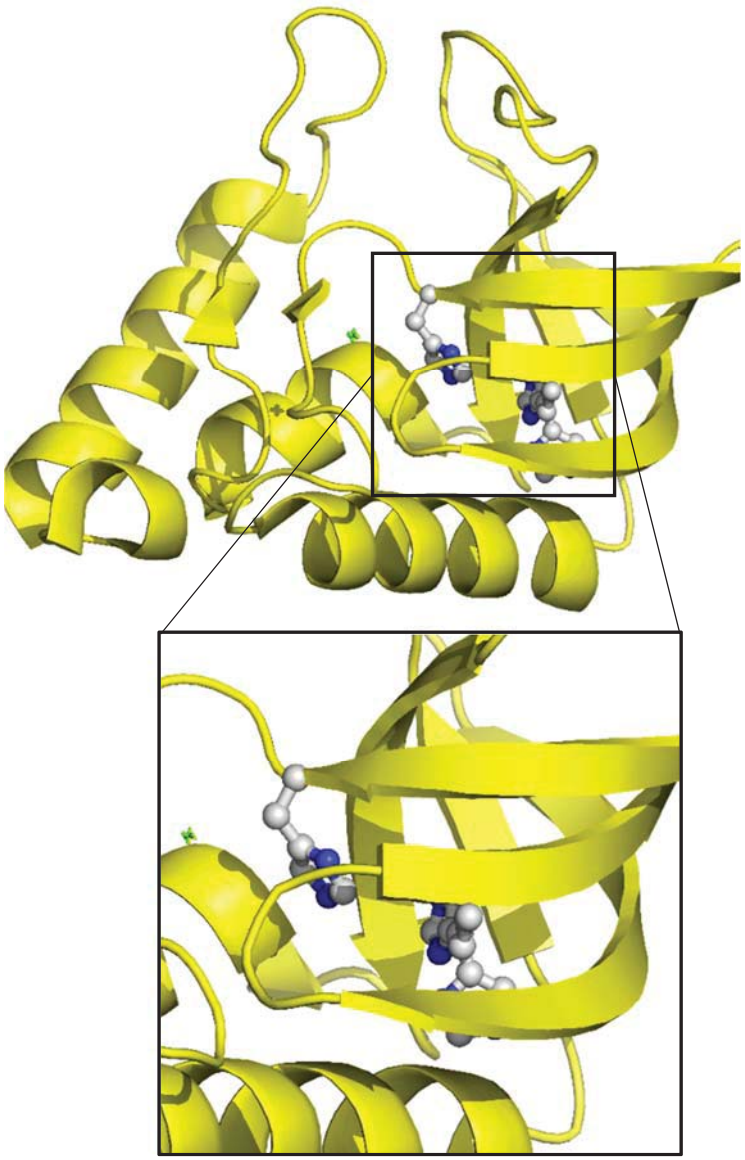

Figure 18: Δ+PHS/V23H/L36H with histidine substitutions at positions 23 and 36. The figure was generated using Pymol [10].

The alignment for the double histidine model protein show that the model is possesses fewer substitutions than previous models (figure 19). The estimated ΔG_{H2O} is 2.4 kcal/mol, which indicated that the protein would be more thermodynamically stable than previous iterations of the protein switch. This would result in a CD spectra more similar to SNase in its native-fold.

		Estimated ΔG°… (kcal/mol)
Reference	Δ+PHS	11.9
Double Histidine	Δ+PHS / V23H / L36H	2.4

Amino acid sequence alignments

Wild-Type	ATSTKKLHKEPATLIKAIDGDTVKLMYKGQPMTFRLLLVDTPETKHPKKGVEKYGPEASAFTKKMVENAKKIEVE
Reference	ATSTKKLHKEPATLIKAIDGDTVKLMYKGQPMTFRKLLVDTPE------**FN**EKYGPEASAFTKKMVENAKKIEVE
Double Histidine	ATSTKKLHKEPATLIKAIDG**D**T**H**KLMYKGQPMTFR**H**LLVDTPE------**FN**EKYGPEASAFTKKMVENAKKIEVE
Wild-Type	FDKGQRTDKYGRGLAYIYADGKMVNEALVRQGLAKVAYVYKPNNTHEQHLRKSEAQAKKEKLNIWSEDNADSGQ
Reference-v2.0	FDKGQRTDKYGRGLAYIYADGKMVNEALVRQGLAKVAYVYK**G**NNTHEQ**L**LRK**A**EAQAKKEKLNIWSEDNADSGQ
Double Histidine	FDKGQRTDKYGRGLAYIYADGKMVNEALVRQGLAKVAYVYK**G**NNTHEQ**L**LRK**A**EAQAKKEKLNIWSEDNADSGQ

Figure 19: Double histidine model and Δ+PHS amino acid sequence alignment with wild-type SNase. Amino acids that have been altered from the wild-type sequence are **bold**. Amino acids changed from the Δ+PHS background are indicated in orange.

Protein was prepared and purified using the established protocol. Circular dichroism spectroscopy was used to assess the secondary structure of Δ+PHS/V23H/L36H. Experiments were performed testing Δ+PHS/V23H/L36H against a positive control that retains the native fold of SNase, Δ+PHS and a negative control that is completely unfolded, WT/T62P. The variants containing single-substituted histidines at positions 23 and 36 were also tested, Δ+PHS/V23H and Δ+PHS/L36H. All proteins were analyzed after incubation in the absence and presence of Ni^{2+} ion over a range of incubation times and at several pH conditions. The spectra in figures 20.A and 20.B suggest that there is no discernible difference between pH 7 and pH 8. This allowed future experiments to be conducted at a singe pH value (pH 8) without likelihood that pH over this narrow range would influence the proteins' secondary structure.

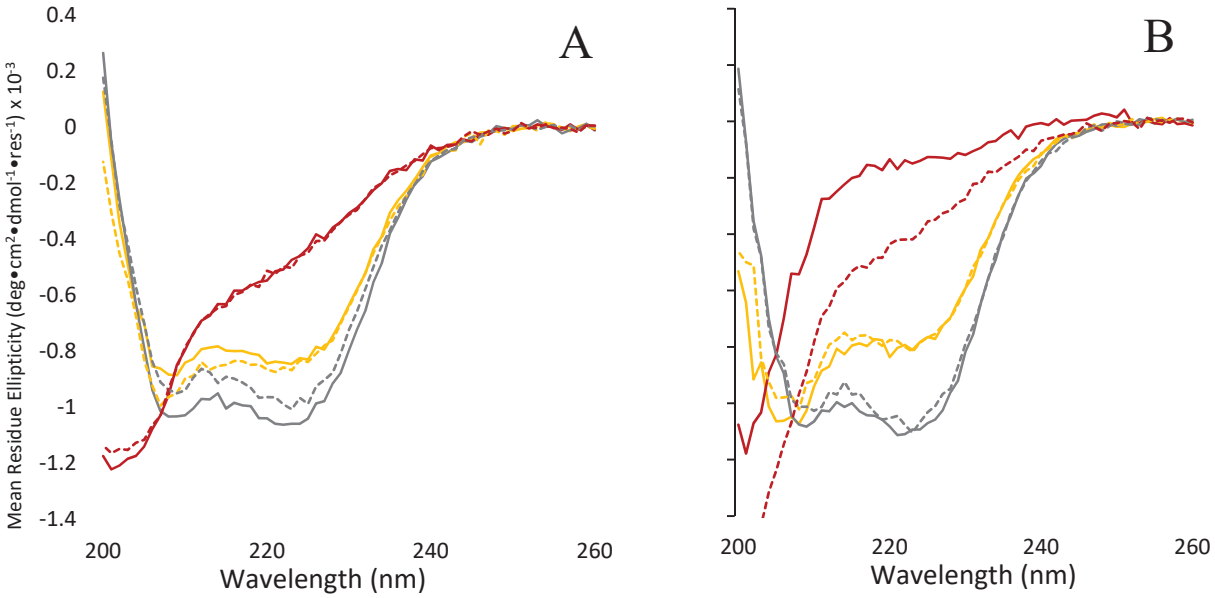

Figure 20.A: Δ+PHS/V23H/L36H (orange, solid, pH7; orange, dotted, pH8) at different pH levels without metal ion retains most of its native fold when compared to both a positive and negative control (Δ+PHS, grey, solid, pH7; grey, dotted, pH8; and WT/T62P, red, solid, pH7; red, dotted, pH8, respectively). There is no noticeable pH effect in either the candidate protein or negative control and a slight depression of the Δ+PHS variant at pH7. **20.B:** Δ+PHS/V23H/L36H (orange, solid, pH7; orange, dotted, pH8) at different pH levels without metal ion retains most of its native fold when compared to both a positive and negative control (Δ+PHS, grey, solid, pH7; grey, dotted, pH8; and WT/T62P, red, solid, pH7; red, dotted, pH8, respectively). There is no noticeable pH effect in either the candidate protein or positive control. The WT/T62P variant curve at pH7 has been translated vertically (red).

Short incubations of the protein in the presence of Ni^{2+} for less than one hour immediately prior to testing the protein sample suggested that there may be a modest recovery of secondary structure for Δ+PHS/V23H/L36. In figure 21.A, the Δ+PHS/V23H/L36H incubated with the metal ion (solid, orange) had a lower peak at 222nm such that it more closely resembles the Δ+PHS positive control with and without the metal ion incubation. Meanwhile, the candidate protein without metal ion incubation had a less pronounced peak, which resembled the spectra of the unfolded negative control. Additional data collection and statistical analyses will be required to establish whether these differences are significant.

To determine if this effect was observable after a longer incubation time, the experiment was repeated using overnight incubation of the proteins in the presence of excess Ni^{2+}, with shaking at 220 rpm at 6°C. Under these conditions, Δ+PHS/V23H/L36H showed no marked difference between protein incubated with metal ion and protein incubated without (figure 21.B). It is possible that, given enough incubation time, the protein will naturally recover a native-fold state. This could be due to the positive ΔG_{H2O} value of Δ+PHS/V23H/L25H, which suggests that the protein will tend more towards native-fold. Further testing will be required to confirm this hypothesis.

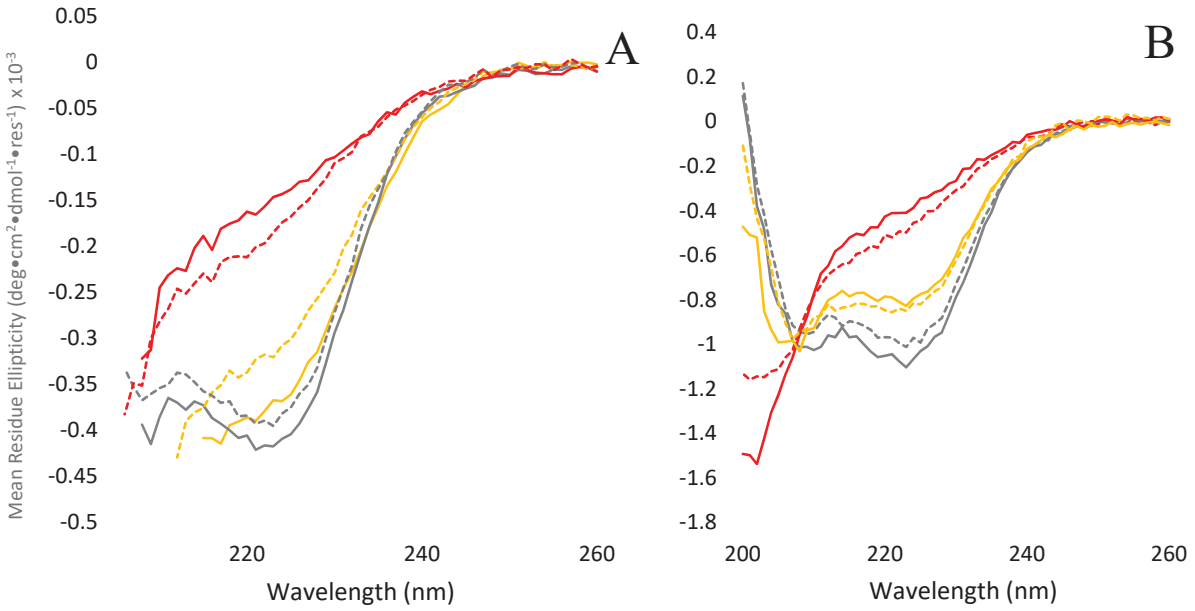

Figure 21.A: Δ+PHS/V23H/L36H at pH 8 with and without metal ion depicted with both a positive and negative control (Δ+PHS, grey, dotted, without Ni^{2+}; grey, solid, with Ni^{2+}; and WT/T62P, red, dotted, without Ni^{2+}; red, solid, with Ni^{2+}, respectively) with incubation occurring immediately prior to data collection. The spectra of Δ+PHS/V23H/L36H without Ni^{2+} (orange, dotted) is slightly raised with a less distinct peak when compared to candidate protein with metal ion (orange, solid), indicating that immediate incubation with Ni^{2+} may result in some recovery of native fold. 21**.B:** Δ+PHS/V23H/L36H at pH 8 with (orange, solid) and without (orange, dotted) metal ion depicted with both a positive and negative control (Δ+PHS, grey, dotted, without Ni^{2+}; grey, solid, with Ni^{2+}; and WT/T62P, red, dotted, without Ni^{2+}; red, solid, with Ni^{2+}, respectively) incubated overnight with shaking at 220rpm at 6°C. There is no discernible difference between the spectra for candidate protein incubated in the absence and presence of the metal ion, indicating overnight incubation with Ni^{2+} has no effect on protein structure.

Discussion and Future Plans

In this protein engineering study, staphylococcal nuclease was modified in an attempt to create a metal ion-dependent switch that would be unfolded in the absence of a metal ion and folded in its presence. Amino acid sequence substitutions were identified in SNase to attempt to approximate the geometry of the high-affinity Ni^{2+} binding site in the metalloregulatory protein NikR. These amino acid substitutions were directed by nucleotide changes introduced into the SNase gene using molecular biology techniques. Two reference proteins and four candidate Ni^{2+}-binding proteins were produced in bacterial host cells, purified, and characterized using gel electrophoresis and circular dichroism (CD) spectroscopy. Denaturing gel electrophoresis in the absence of reducing agent revealed a single band at 16 kDa for both reference proteins, indicating highly purified protein of a size consistent with SNase, with bands at both 16 kDa and 32 kDa for both claw-v1.0 and claw-v2.0 proteins. The double band suggests that a mixed population of monomeric and dimeric proteins may have existed. Parallel gel electrophoresis in the presence of the reducing agent β-mercaptoethanol yielded a single band at 16 kDa for each (reference proteins and claw proteins). Because wild-type SNase and the background proteins do not contain any cysteine residues, and the claw-v1.0 and claw-v2.0 proteins each contain a single cysteine residue, it seems likely that under oxidizing conditions (in air, in the absence of reducing agent), two neighboring unfolded claw-1.0 (or claw-v2.0) proteins formed a dimer through a disulfide bond. Gel electrophoresis indicated that reverting position 34 from cysteine to phenylalanine in claw-v3.0 successfully reduced the occurrence of disulfide bonds, as the protein migrated as a single electrophoretic band at 16 kDa in the absence and presence of reducing agent.

Both reference proteins maintained a native-like fold based on CD spectra. The "W"-shaped spectra with negative peaks at 222 nm and 208 nm that are characteristic of alpha helical proteins were observed for both reference proteins, consistent with other SNase proteins known to adopt a native fold. In contrast, CD spectra indicated that claw-v1.0, claw-v2.0 and claw-v3.0 proteins were partially unfolded, with a loss of signal at 222 nm and the gain of a broad negative peak near 195 nm, consistent with a lack of extensive secondary structure in the proteins. Incubation of reference-v1.0 and reference-v2.0 proteins with a 1000-fold molar excess of Ni^{2+} had no effect on protein secondary structure, as expected. The analogous treatment of claw-v1.0, claw-v2.0 and claw-v3.0, unfortunately, failed to induce a native-fold, as no change in the CD spectra was observed for these proteins. Additional CD experiments in which the reducing agent β-mercaptoethanol was also included with the protein, buffer, and Ni^{2+} failed to induce protein folding for claw-v1.0. The experiment was not repeated for claw-v2.0 or claw-v3.0. Protein aggregation was not likely responsible for the protein failing to return to its native fold, as the flash dilution of claw-v3.0 in Ni^{2+} did not induce protein folding. Likewise, there was not likely an intermediate fold that was thermodynamically stable enough to prevent the protein from binding to Ni^{2+} given that the dialysis in denaturant with Ni^{2+} of claw-v3.0 failed to return the protein to its native fold.

In an attempt to find the point at which the thermodynamic penalty was so great that structure recovery was no longer possible, a candidate containing two histidine substitutions was produced. This candidate appeared to retain most of its native-fold even in the absence of metal ion. Because the spectra for Δ+PHS/V23H/L36H both in the absence and presence of Ni^{2+} suggests that the protein is either a mixture of partially folded and native-folded protein or the protein has lost some of it's α-helical structure, the smallest amount of chemical denaturant

should unfold the protein. This study should be conducted to confirm the fold. Circular dichroism spectra indicate that there is no significant pH effect in the folding of this protein. It is possible that incubation time plays a role in recovery of structure. The peak at 222nm for Δ+PHS/V23H/L36H with Ni^{2+} ion incubation more closely resembled the peaks of the natively folded protein, Δ+PHS with and without Ni^{2+}. However, when shaking incubation overnight occurred on the same protein candidate, the data displayed similar spectra for the candidate protein's secondary structure both with and without metal ion incubation. A possible explanation for this is that the more time over which incubation occurred, the greater the possibility that the protein became mechanically unfolded enough that both samples were able to re-fold into the same thermodynamic minimum, leading to a closer similarity in their CD spectra when compared to the spectra of Δ+PHS/V23H/L36H in the presence and absence of Ni^{2+} collected immediately after incubation. Further testing controlling this incubation time variable will provide more insight.

Currently, it is unclear which aspect of the Ni^{2+}-binding site prevented Ni^{2+} uptake and protein folding. It is possible that a different incubation protocol is required, or a different pH value to optimize protonation state of the introduced histidine side chains. It is also possible that the binding site, about 50% larger in volume than that of the NikR site, is too large for effective coordination of Ni^{2+} by the protein. Alternatively, the thermodynamic penalty introduced by the destabilizing amino acid substitutions in claw-v1.0, claw-v2.0 and claw-v3.0 are too devastating to be rescued by binding of a single Ni^{2+} to the protein.

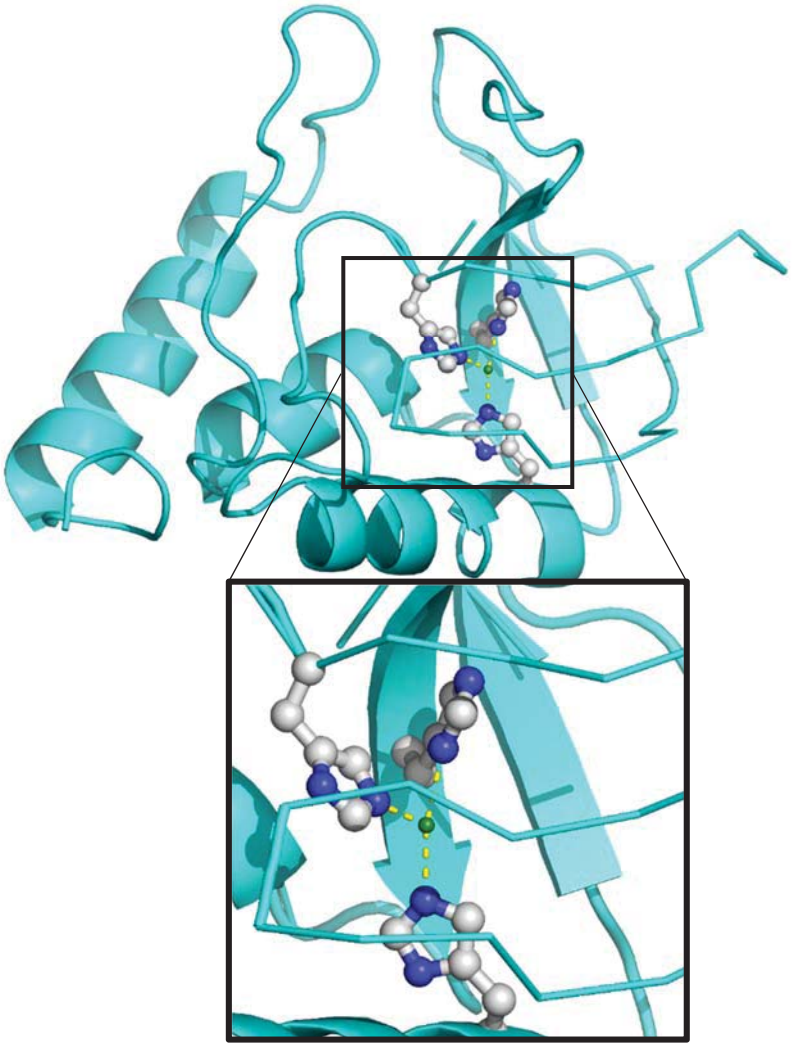

Figure 22: Ribbon diagram of a claw-v4.0 model with the histidine residues depicted at positions 36, 66 and 92. The figure was generated using Pymol [10].

To test several of these hypotheses, work is underway to create the gene construct for claw-v4.0 (figure 22), which will shift the binding site to a separate location in the SNase protein, in an effort to minimize the thermodynamic penalties of the amino acid substitutions. This protein will retain the substitution of alanine at position 23 to minimize steric constraints near the binding site. It will revert the histidine at position 25 to leucine, retain the histidine substitutions at positions 36 and 66, and will substitute a histidine at position 92 (in lieu of substituting an alanine at position 92). This could reduce the thermodynamic penalty by reducing the number of amino acid substitutions necessary to form the binding site. The proposed site is also predicted to be smaller, with binding distances from histidine to nickel ranging from 2.0-2.4 Å, similar to the distances between side chains and the nickel(II) ion in the NikR high-affinity binding site. However, the square planar geometry in the NikR binding site has been altered in claw-v4.0.

Other work will continue to add histidine substitutions iteratively to the background protein. Previous studies with single histidine substitutions at the proposed positions revealed proteins with native-like folds. This approach will allow for assessment of the protein's relative thermostability as each successive amino acid substitution is introduced. The initial metal switch protein designs were based on the square planar Ni^{2+} binding site of NikR. Amino acids were systematically removed to the minimum number necessary for the binding site. Following preliminary work in this project creating a protein with two histidine ligands, a third histidine ligand could be added in the vicinity of the existing two ligands. Alternatively, other variants containing two histidine substitutions in a different site in the protein could be created. By studying such variants, thermodynamic stability can be assessed for each protein as well as Ni^{2+} binding affinity of the site. These experiments utilize the hyperstabilized background Δ+PHS,

which is slightly less stable than Δ+NVIAGLA but is better-characterized. Several future variants include Δ+PHS/L25H/L36H, Δ+PHS/L36H/V66H, Δ+PHS/V66H/I92H and Δ+PHS/V23H/V66H. These candidates were modeled in Pymol [10] and selected for the thermodynamic penalties those substitutions would impose upon the protein and for the geometry of the resulting putative binding sites. After these two-substitution mutations have been characterized, three-substitution proteins could be considered.

Other future work will include incubation with larger metal ions. The larger ligand in the square-planar binding site may aid in the recovery of the native-fold of claw proteins (table 1). Palladium (II) and platinum (II) are both likely d^8 candidates with larger ionic radii than nickel (II), so they may better fit in the proposed SNase internal binding site. Solubility limitations for these larger ions in aqueous conditions would need to be overcome prior to testing.

Table 1. Candidate metal ion for binding to SNase metal ion binding candidates.

Metal Ion	**Ionic Radius (pm)**	**Coordination**
Nickel (II)	**63**	**d^8**
Palladium (II)	**78**	**d^8**
Platinum (II)	**74**	**d^8**

Protein engineering, particularly involving protein-metal ion interactions, remains a formidable challenge. This project created a series of protein variants with multiple amino acid substitutions relative to the candidate's background protein. The endeavor was to engineer a "switch" protein in an attempt to understand better the molecular underpinnings of metalloprotein folding. While a fully-functional protein switch candidate was not obtained, iterative steps toward this goal were achieved. Partially unfolded proteins were produced and characterized. These approaches are essential due to the limitations of current computational

methods for metalloprotein folding, which remain in development.

The design and synthesis of four protein candidates has led to contributions in understanding protein engineering. The formation of a disulfide bond in claw-v1.0 and claw-v2.0 underscores the fact that are challenges in substituting a cysteine residue into the interior of a protein which should be considered carefully in future protein candidates. Other considerations of note are the effects of steric constraints and thermodynamic penalties. In the candidates generated, proteins with four to six amino acid substitutions were unfolded. This result was not surprising upon consideration of the proteins' estimated thermodynamic instabilities, calculated using single-variant empirical data. In contrast, the protein with two amino acid substitutions retained a majority of its native-fold, consistent with its estimated thermodynamic stability (2.4 kcal/mol). A successful protein switch candidate, capable of binding to Ni^{2+} and refolding, likely will contain between two and four amino acid substitutions. The current study, therefore, has narrowed the scope of study for future experiments.

References

1. Nelson, David L., David L. Nelson, Albert L. Lehninger, and Michael M. Cox. *Lehninger Principles of Biochemistry*. New York: W.H. Freeman, 2008.

2. Wright, C. M., Heins, R. A., and Ostermeier, M., "As Easy as Flipping a Switch?" *Current Opinion in Chemical Biology* 2007, 11:342–346.

3. Stratton, M. M. and Loh, S. N., "Converting a protein into a switch for biosensing and functional regulation," *Protein Science* 2010, 20:19–29.

4. D. Shortle and A.K. Meeker, Biochemistry, 28 (1989) 936.

5. Robinson, A.C., Castañeda, C.A., Schlessman, J.L., and García-Moreno, B., "Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein," *PNAS* 2014, 111:11685–11690.

6. Chivers, P. T. and Tahirov, T. H., "Structure of Pyrococcus horikoshii NikR: Nickel Sensing and Implications for the Regulation of DNA Recognition," *Journal of Molecular Biology* 2005, 348:597–607.

7. Li, Wenfei, Wang, Jun, Zhang, Jian, and Wang, Wei. "Molecular simulations of metal-coupled protein folding." *Science Direct* 2015, 30:25-31.

8. Zambelli, B., Bellucci, M., Danielli, A., Scarlato, V., and Ciurli, S., "The Ni^{2+} binding properties of Helicobacter pylori NikR," *Royal Society of Chemistry* 2007, 3649–3651.

9. Benini, S., Cianci, M., and Ciurli, S., "Holo-Ni^{2+}-Helicobacter pylori NikR contains four square-planar nickel-binding sites at physiological pH," *Royal Society of Chemistry* 2011, 20:7831–7833.

10. PyMOL, accessed 7 January 2016 at http://pymol.org/.

11. Robinson, A.C., Majumdar, A., Schlessman, J.L., Garcia-Moreno E., B. "Charges in Hydrophobic Environments: A Strategy for Identifying Alternative States in Proteins" *Biochemistry* 56(1), 212-218, 2017.

12. Roche, Julien, Caro, Jose A., Norberto, Douglas R., Barthe, Phillipe, Roumestand, Christian, Schlessman, Jamie L., Garcia, Angel E., García-Moreno, Bertrand E., Royer, Catherine. "Cavities determine the pressure unfolding of proteins." *PNAS* 18:109 6945-6950. May 1, 2012. Web.

13. Marbach, Anja, and Katja Bettenbrock. "Lac Operon Induction in *Escherichia coli*: Systematic Comparison of IPTG and TMG Induction and Influence of the Transacetylase LacA." *Journal of Biotechnology* 157.1 (2012): 82-88. Web.

14. Barr, Frederic G., Michael B. Kastan, and Michael W. Lieberman. "Distribution of 5-methyldeoxycytidine in Products of Staphylococcal Nuclease Digestion of Nuclei and Purified DNA." *Biochemistry* 24.6 (1985): 1424-428. Web.

15. Greenfield, N. "Using circular dichroism spectra to estimate protein secondary structure," *Nature Protocols* 2006, 1(6): 2876-2890.

16. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., McNicholas., S.J., Murshudov, G.N., Pannu, N.S., Potterton, P.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67 (Pt 4): 235-242.

17. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., Read, R.J. (2005) Likelihood-enhanced fast translation functions. Acta Crystallogr D Biol Crystallogr 61(Pt 4): 458–464.

18. Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholis, R.A., Winn, M.D., Long, F., Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67 (Pt 4): 355-367.

19. Emsley, P., Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60 (Pt 12 Pt 1): 2126–2132.

Bibliography

Baran, Kelli L. Chimenti, M.S., Schlessman, J.L, Herbst, K.J., and García-Moreno, B.E. "Electrostatic Effects in a Network of Polar and Ionizable Groups in Staphylococcal Nuclease." *Journal of Molecular Biology*, vol. 379, no. 5, 2008, pp. 1045–1062.

Barr, Frederic G., Michael B. Kastan, and Michael W. Lieberman. "Distribution of 5-methyldeoxycytidine in Products of Staphylococcal Nuclease Digestion of Nuclei and Purified DNA." *Biochemistry* 24.6 (1985): 1424-428. Web.

Benini, S., Cianci, M., and Ciurli, S., "Holo-Ni^{2+}-Helicobacter pylori NikR contains four square-planar nickel-binding sites at physiological pH," *Royal Society of Chemistry* 2011, 20:7831–7833.

Castaneda, Carlos A. Fitch, C.A., Majumdar, A., Khangulov, V., Schlessman, J.L., and García-Moreno, B.E. "Molecular Determinants of the pK_a values of Asp and Glu Residues in Staphylococcal Nuclease." *Proteins: Structure, Function, and Bioinformatics*, vol. 77, no. 3, 2009, pp. 570–588.

Chivers, P. T. and Tahirov, T. H., "Structure of Pyrococcus horikoshii NikR: Nickel Sensing and Implications for the Regulation of DNA Recognition," *Journal of Molecular Biology* 2005, 348:597–607.

D. Shortle and A.K. Meeker, Biochemistry, 28 (1989) 936.

Denisov, Vladimir P., Schlessman, Jamie L., García-Moreno, Bertrand E., and Halle, Bertil. "Stabilization of Internal Charges in a Protein: Water Penetration or Conformational Change?" *Biophysical Journal*, vol. 87, no. 6, 2004, pp. 3982–3994.

Emsley, P., Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60 (Pt 12 Pt 1): 2126–2132.

Greenfield, N. “Using circular dichroism spectra to estimate protein secondary structure,” Nature Protocols 2006, 1(6): 2876-2890.

Li, Wenfei, Wang, Jun, Zhang, Jian, and Wang, Wei. “Molecular simulations of metal-coupled protein folding.” *Science Direct* 2015, 30:25-31.

Marbach, Anja, and Katja Bettenbrock. "Lac Operon Induction in *Escherichia coli*: Systematic Comparison of IPTG and TMG Induction and Influence of the Transacetylase LacA." *Journal of Biotechnology* 157.1 (2012): 82-88. Web.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., Read, R.J. (2005) Likelihood-enhanced fast translation functions. Acta Crystallogr D Biol Crystallogr 61(Pt 4): 458–464.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholis, R.A., Winn, M.D., Long, F., Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67 (Pt 4): 355-367.

Nelson, David L., David L. Nelson, Albert L. Lehninger, and Michael M. Cox. *Lehninger Principles of Biochemistry*. New York: W.H. Freeman, 2008.

Passerini, A. Lippi, M., and Frasconi, P. “MetalDetector v2.0: Predicting the Geometry of Metal Binding Sites from Protein Sequence.” *Nucleic Acids Research*, vol. 39, 2011.

Perrin, D. D., and I. G. Sayce. “Complex Formation by Nickel and Zinc with Penicillamine and Cysteine.” *Journal of the Chemical Society A: Inorganic, Physical, Theoretical*, 1968, p. 53.

PyMOL, accessed 7 January 2016 at http://pymol.org/.

Regan, Lynne. “Protein Design: Novel Metal-Binding Sites.” *Trends in Biochemical Sciences*, vol. 20, no. 7, 1995, pp. 280–285.

Robinson, A. C. Castaneda, C.A., Schlessman, J.L., and García-Moreno, E.B. “Structural and

Thermodynamic Consequences of Burial of an Artificial Ion Pair in the Hydrophobic Interior of a Protein." *Proceedings of the National Academy of Sciences*, vol. 111, no. 32, 2014, pp. 11685–11690.

Robinson, A.C., Majumdar, A., Schlessman, J.L., Garcia-Moreno E., B. "Charges in Hydrophobic Environments: A Strategy for Identifying Alternative States in Proteins" *Biochemistry* 56(1), 212-218, 2017.

Robinson, A.C., Castañeda, C.A., Schlessman, J.L., and García-Moreno, B., "Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein," *PNAS* 2014, 111:11685–11690.

Roche, Julien, Caro, Jose A., Norberto, Douglas R., Barthe, Phillipe, Roumestand, Christian, Schlessman, Jamie L., Garcia, Angel E., García-Moreno, Bertrand E., Royer, Catherine. "Cavities determine the pressure unfolding of proteins." *PNAS* 18:109 6945-6950. May 1, 2012. Web.

Schlessman, Jamie L. Abe, Colby, Gittis, Apostolos, Karp, Daniel A., Dolan, Michael A., and García-Moreno, Bertrand E. "Crystallographic Study of Hydration of an Internal Cavity in Engineered Proteins with Buried Polar or Ionizable Groups." *Biophysical Journal*, vol. 94, no. 8, 2008, pp. 3208–3216.

Stratton, M. M. and Loh, S. N., "Converting a protein into a switch for biosensing and functional regulation," *Protein Science* 2010, 20:19–29.

Ullmann, R. Thomas, and G. Matthias Ullmann. "GMCT : A Monte Carlo Simulation Package for Macromolecular Receptors." *Journal of Computational Chemistry*, vol. 33, no. 8, 2012, pp. 887–900.

Wright, C. M., Heins, R. A., and Ostermeier, M., "As Easy as Flipping a Switch?" *Current Opinion in Chemical Biology* 2007, 11:342–346.

Zambelli, B., Bellucci, M., Danielli, A., Scarlato, V., and Ciurli, S., "The Ni^{2+} binding properties of Helicobacter pylori NikR," *Royal Society of Chemistry* 2007, 3649–3651.

Appendix I: Glossary

$\Delta G°_{H2O}$: Gibbs free energy, a thermodynamic quantity to describe stability. By this convention, negative values are unstable and positive values are stable.

$\Delta\Delta G°_{H2O}$: The thermodynamic penalty for a single amino acid substitution at a specific location.

Aliquot: a portion of a whole sample.

Chemical denaturant: chemical agents that induce unfolding in proteins.

Chiral: molecule structured in such a way that the structure and its mirror image are not superimposable.

Cysteine: amino acid with a thiol side-chain capable of undergoing oxidation and is thus susceptible of bonding to other thiol groups to form a disulfide bridge.

Histidine: a positively charged amino acid at physiological pH possessing an imidazole side-chain.

Imidazole: an organic compound formed by an aromatic ring possessing nitrogen atoms at positions 1 and 3, carbon atoms at positions 2, 4, and 5, and a double bond between positions 2 and 3 and 4 and 5.

Metalloprotein: a protein with a metal ion cofactor.

Monomeric: describing single-unit polypeptide chains.

Oligomerization: the formation of complex molecules from simpler subunits.

Oligonucleotide: short DNA or RNA molecules.

RNA polymerase: enzyme necessary to transcript RNA chains from DNA.

Staphylococcal nuclease (SNase): a small protein used for cleaving DNA or RNA substrates that has served as a model system for protein folding studies.

Steric constraints: effects caused by the way molecules occupy space in three dimensions, which limits potential chemical reactions.

Supernatant: The liquid that has been separated from solid material after centrifugation.

Thermodynamic penalty: a consequence to overall protein stability due to the substitution of an amino acid not native to the protein's sequence.

Thiol: a group with a carbon-bonded sulfur.

Appendix II: Index of Figures